

# 面向英文新冠推文的实体 情感标注指南<sup>1</sup>

YLab  
浙江大学  
版本： 1.0

---

<sup>1</sup>本规范的制定参考自 Orbach M, Toledo-Ronen O, Spector A, et al. YASO: A Targeted Sentiment Analysis Evaluation Dataset for Open-Domain Reviews[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021: 9154-9173

1. 背景介绍：	3
2. 标注原则：	3
3. 具体情况：	4
3.1 正面及负面情感判断方式：	4
直接判断：	4
直接表达：	4
间接表达：	5
3.2 基于不同分类的特殊案例标注规则：	6
<i>Drug 和 Vaccine:</i>	6
<i>Person 和 Organization:</i>	6
4. 标注示例：	7
4.1 规则示例：	7
4.2 标注错误示例	8

## 1. 背景介绍:

新冠流行期间,用户对不同实体(如人物、机构、疫苗、药物等)的态度激烈变化。利用社交媒体大数据,实时追踪用户对不同实体的态度变化有助于疾控部门制定疫情政策、宣传疫苗接种的必要性、减少错误信息传播等。本标注指南针对新冠大流行期间的重点实体类别(人物、机构、疫苗、药物)进行用户情感标注,构建目标级情感(Targeted Sentiment)数据库,即针对目标实体进行情感标注。不同于句子级别的情感标注,Targeted 情感分析是一种更细粒度的情感分析方法,能够精细地区分用户对具体实体的情感态度,而非对整个句子进行粗略情感判断,因此更加具体和有更高的利用价值。

首先将新冠病毒大流行期间的 twitter 数据进行实体标注,并且将该数据作为实体情感标注的源数据。在实体情感标注中,本指南主要关注如下几大实体内部蕴含的情感:药物(Drug)、疫苗(Vaccine)、人物(Person)、组织(Organization)。研究人员可以使用实体情感数据集用于新冠肺炎大流行期间社交媒体用户对不同实体的情绪和态度变化分析。

## 2. 标注原则:

将用户对目标实体的情感标注为正,负,中性三大类型。

- A. Twitter 整体考虑: 标注过程中需要基于整条 Twitter 内容综合判断用户对实体的情感。不能简单考虑单句话对实体的影响,需要通读整条 Twitter 后再进行判断。
- B. 用户角度: 在标注过程中需要从用户的角度出发分析实体情感,判断 Twitter 用户本人对 Twitter 文本中实体的情感。
- C. 反讽: 如果能够明确文本是有反讽的意味,则应该按照用户反讽的真实意思进行标注。
- D. 转述: 针对由用户转述的 twitter, 需要以用户的角度判断情感,判断用户的情感。如无法判断用户的态度,则标注为中性。

- E. 关注特殊符号/格式/反问：分析情感的过程中要结合 twitter 的标点如感叹号、引号、大写、emoji 文字等，使用这些特殊的书写方式一般情感更强烈，是情感标注的重要线索。
- F. 缺乏上下文信息：有一些 tweet 缺乏上下文，难以标注其内部实体情感，不需要过度深究，标注中性即可

### 3. 具体情况：

#### 3.1 正面及负面情感判断方式：

直接判断：

直接表达：

1. 用户对实体直接表达情感，①带有情感色彩的词直接形容实体②带有情感地评价实体好/坏的影响③描述实体产生的积极/消极的行为④推荐/不推荐或接受/拒绝
2. Twitter 中含有反讽的意味，通常为负面情感，判断方式为看似积极的描述+不良行为/后果
3. Tweet 中有对某个实体直接的情感态度，基本上按照该情感态度对实体进行标注。

案例：

- `[@Sean Hannity#Person^Negative*]` is a biased propaganda spewing douche bag.
- `[@Trump#Person^Negative*]`'s erratic trade and foreign policy creates big headaches for businesses going forward. Add that to good chance of COVID spike w/ `[@flu#Disease*]` season killing more people and jobs. Not much for market 2 cheer...
- Dr gave her `[@ivermectin#Drug^Positive*]`. 2 days later up and about, no `[@fever#Symptom*]`, `[@cough#Symptom*]` gone. Still `[@tired#Symptom*]` but otherwise much better. I would take it if diagnosed positive for covid.

- If we stopped testing now, we'd have very few cases (of corona) says  
[@trump#Person^Negative\*] Almost directly from "The House of  
God" "You can't find a [@fever#Symptom\*] if you dont take a  
temperature" Excellent book, btw, which everyone should read

间接表达:

1. 根据用户对某一个实体的情感对另一实体的情感进行间接判断，  
如 A 实体与 B 实体为因果关系，则用户对 A 实体和 B 实体的  
关系应一致
2. A 实体与 B 实体为转折/对立关系，则用户对 A 实体与 B 实体的  
情感态度相反
3. 实体一开始有强烈的情感色彩，由于情感存在延续性，后续的相  
同实体可以标注为同一种情感。但是有合理的原因的情况下，实  
体情感可以出现转折。
4. 实体一开始没有强烈的情感色彩，后续存在的对实体的情感不会  
影响一开始的中性态度。
5. 在表述 A 实体时，举出 B 实体作为例子，此时 A 实体和 B 实体  
情感应基本一致

案例:

- Numbers are under reported because [@GOP#Organization^Negative\*]  
goons like [@Ron DeSantis#Person^Negative\*] classify deaths as  
[@pneumonia#Disease\*] rather than Covid, when statistically it's a  
clear cover up for his ineptitude... nice try though, Trumpanzee.
- The data is quite clear. Suburban women are voting  
[@Democratic#Organization^Positive\*]. Apparently they are tired of  
[@Trump#Person^Negative\*]'s whining and his to unending Insults  
but mostly because of his malfeasance regarding COVID 19.
- Numbers are under reported because  
[@GOP#Organization^Negative\*] goons like [@Ron  
DeSantis#Person^Negative\*] classify deaths as

[@pneumonia#Disease\*] rather than Covid, when statistically it's a clear cover up for his ineptitude... nice try though, Trumpanzee.

### 3.2 基于不同分类的特殊案例标注规则：

#### Drug 和 Vaccine:

类型	判断依据	情感
涉及副作用	A. 仅描述副作用，没有其他评价 B. 描述副作用，但后续提到影响不大 C. 描述副作用比较轻微 D. 客观的描述副作用	中性
	A.认为疫苗副作用造成严重影响或疫苗无效 B.夸张的形容副作用	负面
	A.有副作用，认为疫苗可以给自己带来保护/有正面作用 B.有副作用，后续有积极情感表达 C.推荐别人接种	正面

#### Person 和 Organization:

类型	判断依据	情感
政治有关	A. 明确的指出分析支持什么、反对什么 B. 人/组织出现与常理相悖的形容或者点评-CDC 政治化，Trump 的信任鸿沟，不客观的组织，无用的药物	A. 支持-正面 反对-负面 B. 负面
反讽	A. 判断语气	负面

COVID 关注度	A.只关注 COVID 忽略了对其他疾病 B.夸大 COVID 后果	负面
-----------	---------------------------------------	----

## 6. 标注示例:

### 4.1 规则示例:

Example:

tweet: @catturd2 There are no more talking lies he can say. Covid is getting boring, peeps are tired of *[@Biden#Person^Negative\*]* blaming *[@Trump#Person^Neutral\*]*. *[@Charlotte#Person^Negative\*]* lie is been proven wrong. Always the white supremacy lie .*[@Trump#Person^Neutral\*]*'s taxes, who cares. Military lie proven wrong. That's all *[@Biden#Person^Negative\*]* has. 5 talking points (lies).解释: *@Biden#Person*: 认为 Biden 在撒谎, 根据句意应该是负面的

tweet: y'know what? i am so proud of *[@yoongi#Person^Positive\*]*. to go through years of *[@shoulder pain#Symptom\*]* and then undergo surgery that requires tough temporary adjustments to his normal lifestyle all on top of the frickin pandemic which has already changed enough for every single one of us..

*yoongi#Person:so proud of yoong* 应该是正面的

tweet: @KellyannePolls @JoeBiden Well not everyone can have a *[@White House#Location^Neutral\*]* balcony to go *[@cough#Symptom\*]* on for some of us, external spaces will do when it's w/o Covid. *[@Biden#Person^Negative\*]* always wears masks You're petty. And nice try on invoking *[@Hillary#Person^Positive\*]* ... your favorite opponent. As a woman you should be ashamed of demonizing her.

*White House#Location: 虽然大意是对 Biden 持负面态度, 但是对于 White House 是中性的*

## 4.2 标注错误示例

### Organization

#### Twitter1:

@JusticeTargeted It's like they have more time than ever since "other" non Covid related surgeries and other medical appointments are canceled and visitors aren't allowed..So much free time to dance and choreography [@tik tok#Organization^Negative\*] videos....oh and there's no [@flu#Disease\*] season or [@pneumonia#Disease\*] to deal with either.

结论: 这里 tik tok 更多像一种工具 应该是 neutral

#### Twitter2:

@britishlivesuk @val\_tigress @FenrirWolf26 Read [@the BTP#Organization^Positive\*] report. He was obnoxious deliberately [@coughing#Symptom\*] at people on the train to induce covid panic. He thought it was funny until [@BTP#Organization^Positive\*] intervened then the mask story emerged. He swore and abused the cop. Smart aleck lost this time.

结论: 只是陈述事实

### Person

#### Twitter1:

@AshaRangappa\_ The mask did a good job hiding [@Chauvin#Person^Negative\*]'s shock and disbelief

结论: 只说明了 Chauvin 的 shock 和 disbelief, 在没有更多的信息前, 应该是 neutral

#### Twitter2:

As CM [@Kejariwal jii#Person^Neutral\*] after getting [@cough#Symptom\*], [@cold#Symptom\*] and [@fever#Symptom\*] became anxious to be tested for Covid, same anxiousness is there in other persons affected by [@cold#Symptom\*] [@cough#Symptom\*] and



[@fever#Symptom\*]. Then why should he stay at home and wait for more symptoms? Mr [@Kejariwal#Person^Neutral\*] please answer.

结论：根据 Then why should he stay at home and wait for more symptoms?有一定的反讽意味，可以理解为对 Kejariwal 做法持负面态度。

### Twitter3:

@RealPDeLuise @c\_harley48 @LouDPhillips @marcosanchezorg @MaurySterling So sorry for your loss [@Peter#Person^Negative\*]  
@c\_harley48 My thoughts prayers are with you everyone who has lost a loved during these difficult times. My brother in laws mom passed away this morning from [@covid pneumonia#Disease\*] complications. She has been in the hospital since Christmas eve. Sad

结论：这里是安慰的意味，是标注错误

### Twitter4:

@AI\_Clayton For 4 years the left screamed  
[@Trump#Person^Negative\*] was a dictator, loved  
[@Putin#Person^Negative\*], etc Yet [@Biden#Person^Negative\*] is FORCING people businesses to comply with [@COVID vax#Vaccine-related^Negative\*], "they're tired of non-compliant", so ppl want to ban unvaxd from jobs, shopping, medical care, etc. Isn't THAT epitome of "Dictator"?

结论：For 4 years the left screamed [@Trump#Person^Negative\*] was a dictator, loved [@Putin#Person^Negative\*]. 可以从后文看出应该是对 Biden 的强迫的做法持负面态度。标注 Negative 是标注错误。

[@COVID vax#Vaccine-related^Negative\*]整体的态度是批判 Biden 强制接种。

对疫苗的态度尚不明确。

### Twitter5:

@kayleighmcenany [@Trump#Person^Negative\*] asked  
[@Wallace#Person^Negative\*] not to mention the number of Covid-19 deaths that he' s responsible for. That' s more than 200,000. TWO

HU[@ND#Location\*]RED THOUSA[@ND#Location\*]. Better to be tired than inept, unqualified, and a liar like  
[@Trump#Person^Negative\*].

结论：Trump 要求 Wallace 不要提及死亡人数，没有直接证据对他表示负面态度

#### Twitter6:

@goldengateblond @billprady My God in so tired of conservatives whining about [@Obama#Person^Negative\*]'s birthday. It's outdoors. I am area with low transmission. Everyone is vaccinated. It's not only within the letter of the law, it's within the current recommendations of  
[@the CDC#Organization^Positive\*].

结论：这里是抱怨 conservatives 对 Obama 生日的抱怨，标注错误。

#### Twitter7:

For the crowd saying the backup [@Clemson OL#Person^Positive\*] shouldn't have practiced Bc he was “symptomatic”, explain to me like you would to a 5 year old, how you'd justify sitting out a player with a negative Covid test and a [@runny nose#Symptom\*] and/or  
[@congestion#Symptom\*].

结论：这里没有太多的对 Clemson OL 的表扬态度，更多的可能是对观众行为的不认可。

#### Twitter8:

@BillyPrempeh I realized with a shock that  
[@CDC#Organization^Negative\*] was politicized when  
[@obama#Person^Neutral\*] got the director to say that we WANT to have  
[@ebola#Disease\*] patients in[@ the USA#Location\*].

结论：因为 Obama 的话让 CDC 出现政治意味。因此 CDC 是负面的话，作为原因，应该也是负面的。

#### Vaccine

##### Twitter1:

As I cope with [@aches#Symptom\*] and [@chills#Symptom\*] after getting my second [@COVID-19 vaccine#Vaccine-related^Positive\*], I

would like to remind you that Spider-Man 2 had the best pop soundtrack ever.

结论：陈述接种事实，标注错误