Guidelines for Annotation Project Task

Sarcasm Detection and Classification for News Headlines

Written by Abdullah Azhar, and Shane Hogan, Matthew Dworkin

Info 259 - April 10, 2024

Introduction

This document is part of a broader Natural Language Processing (NLP) project in which the task is to detect and classify sarcasm in news headlines. Accompanying this document is a dataset containing headlines from HuffPost and The Onion. The headlines have been annotated by humans with labels indicating *whether* a document contains sarcasm and distinguishing *which* type of sarcasm is present. This document outlines the guidelines that the annotators agreed upon before making the annotations.

Label Overview

Whereas news headlines from HuffPost typically are not sarcastic, those from The Onion typically are. Sarcasm broadly fits into one of three categories: satire, irony, and humor. Altogether, the full set of labels used in these annotations are: {none, satire, irony, humor}.

Label Descriptions

None

The headline describes a real event or article of interest that is not playful, jesting, or intentionally humorous. The headline is straightforward and contains no deviation from the literal meaning of the sentence. It likely came from a legitimate news source (HuffPost) as opposed to an entertainment media source (The Onion).

Example: *"Microsoft fights U.S. government over data requests."* This headline fits the category "none" because it contains no sarcastic remarks and plausibly describes a real event.

Satire

The headline uses exaggeration, or ridicule, expose, criticize, or poke fun at one's stupidity or vices. Although not exclusively, satire usually contains an element of political or social commentary and targets a specific group or individual.

Example: *"Steve Bannon mixes discarded climate change report with saliva to build final wall of nest."* This headline fits the category "satire" because it pokes fun at a particular person (Steve Bannon) who is also a political figure, providing imagery of him as a creature and commenting on his careless attitude toward climate change reports.

Additionally, for some examples, the line between satire and humor also depends on the contextual setting of the post. For example, while 'man sentenced to 3 months probation for 17th-degree murder' could be classified as humor, it is worthwhile to analyze this from a political sphere which could potentially be a mockery of the judicial system leading to an element of satire in this post.

Irony

The headline describes the opposite of reality or the opposite of what one would typically expect. These headlines have a clear disjunction between the literal meaning of the words and the intended meaning of the sentence.

Example: *"3822 voted America's favorite pin number."* This headline fits the category "irony" because it is likely made up and is the opposite of what one would expect; it broadcasts information that is typically kept private.

Humor

While all of the sarcastic headlines contain some element of humor, this class serves as a catch-all for those that are sarcastic in nature but otherwise do not fit the guidelines for satire or irony. Distinguishing from the class "none," these headlines are witty, use wordplay, or contain comedic exaggeration to elicit laughter from the reader.

Example: *"Mother ferries 4 more shirt options back to son in gap dressing room."* This headline first the category "humor" because it is likely sarcastic (i.e. not a real news headline) and was

made up to make the reader laugh. While it may poke some fun at mothers, this headline does not constitute satire because its humor is centered more around the absurdity that this mundane everyday event would make a news headline.