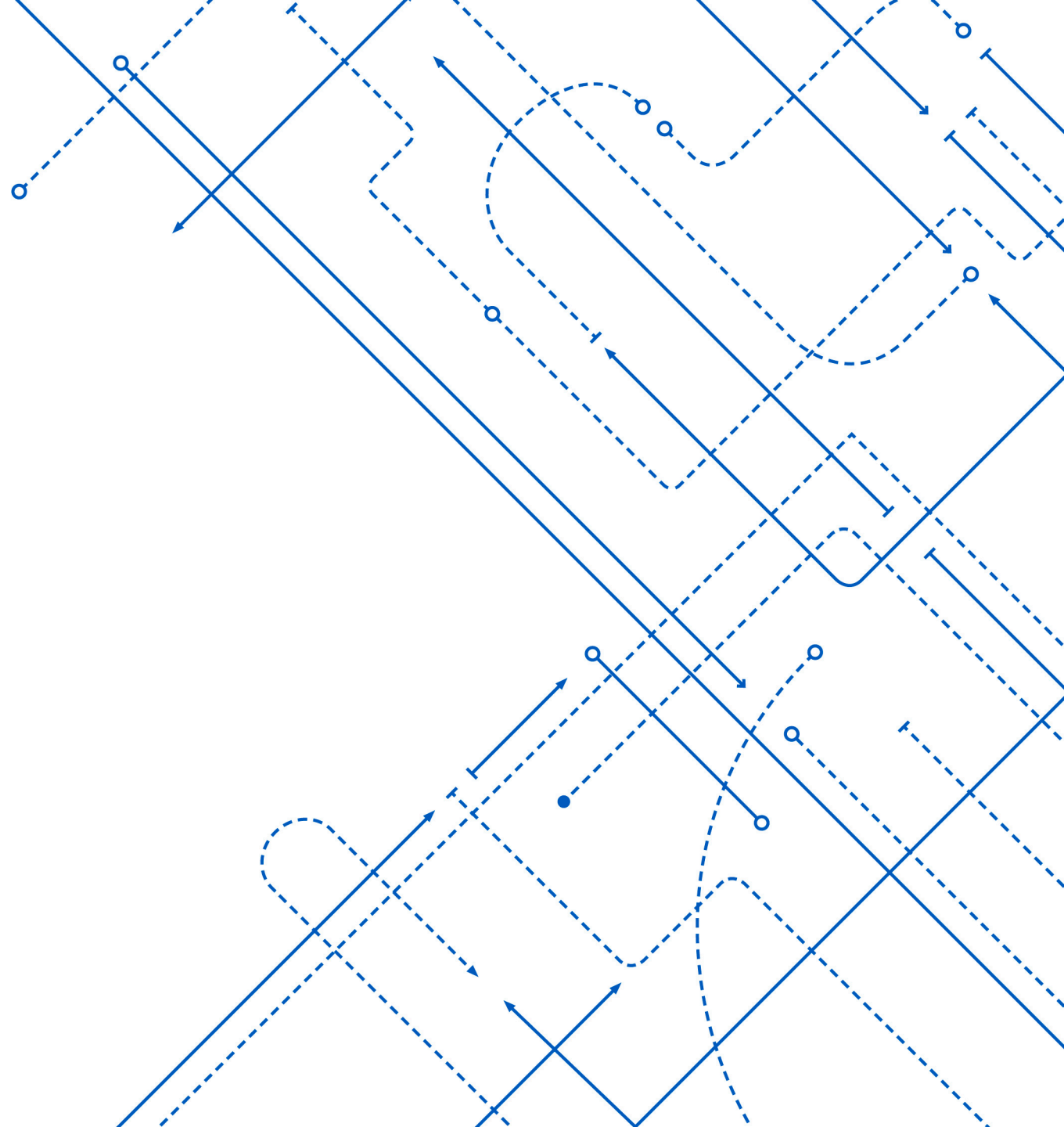
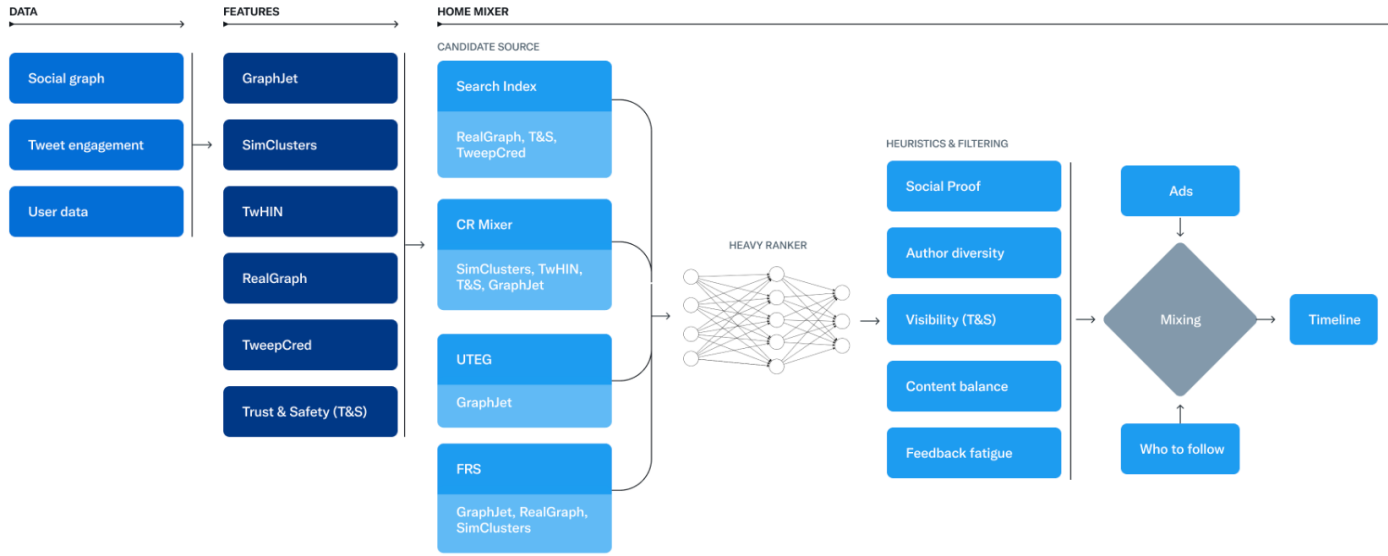
NLP & Social Media



University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

Twitter Recommendation Algorithm

The Twitter Recommendation Algorithm is a set of services and jobs that are responsible for constructing and serving the Home Timeline. For an introduction to how the algorithm works, please refer to our engineering blog. The diagram below illustrates how major services and jobs interconnect.



These are the main components of the Recommendation Algorithm included in this repository:


elon twitter algorithm
Islanders at Hurricanes
(
"author_is_elon",
candidate ⇒
candidate
.getOrElse(AuthorIdFeature, None).contains(candidate.getOrElse(DDGStatsElonFeature, 0L))),
(
"author_is_power_user",
candidate ⇒
candidate
.getOrElse(AuthorIdFeature, None)
.exists(candidate.getOrElse(DDGStatsVitsFeature, Set.empty[Long]).contains)),
(
"author_is_democrat",
candidate ⇒
candidate
.getOrElse(AuthorIdFeature, None)
.exists(candidate.getOrElse(DDGStatsDemocratsFeature, Set.empty[Long]).contains)),
(
"author_is_republican",
candidate ⇒
candidate
.getOrElse(AuthorIdFeature, None)
.exists(candidate.getOrElse(DDGStatsRepublicansFeature, Set.empty[Long]).contains)),
© 2023 Twitter, Inc.

A Weakly Supervised Classifier and Dataset of White Supremacist Language

Anonymous ACL submission

Kenny and Atri's quick summary

- Most people: how would this be used and what were the implications of that?
- Most people liked the idea of weak supervision and the explicit incorporation of anti-racist text
- Common question: Could OpenAI have used something like this? Did they?
- Technical Questions
  - How is text different from images/video?
  - What is a topic / topic model?
  - Various questions about the data

Rest of today / some of Wednesday

- Differentiating various ways of accessing data
- An ad-hoc, mostly question-driven discussion about NLP and evaluation

Where does data come from?

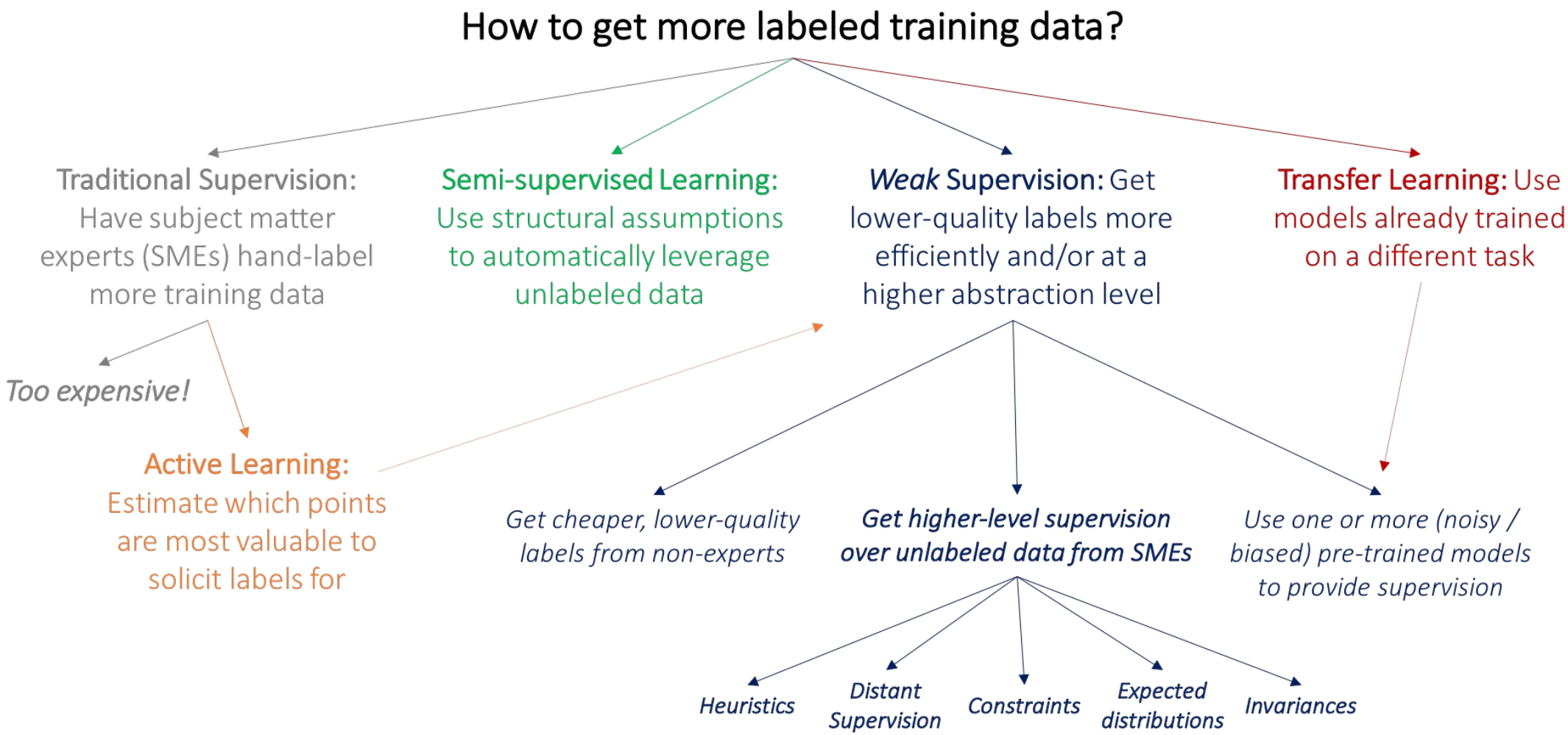
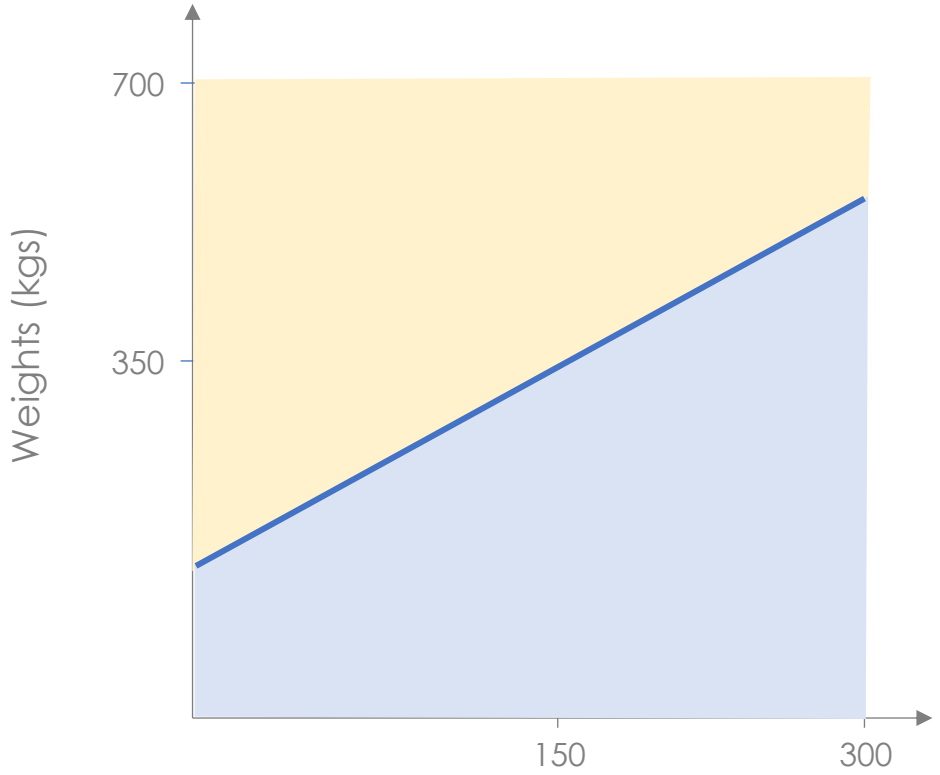


Image source: http://ai.stanford.edu/blog/weak-supervision/

A linear model for classification

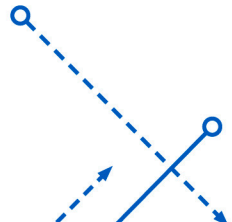


Linear models

The models we considered above (as we have seen before) are called *linear models* (because you can literally draw a line to present them in 2D). In particular, in the case of two input variables, a linear model is **completely defined** once you specify the line as which of the two sides is the positive side (and the other side automatically becomes the negative side).

Sample task: stance detection

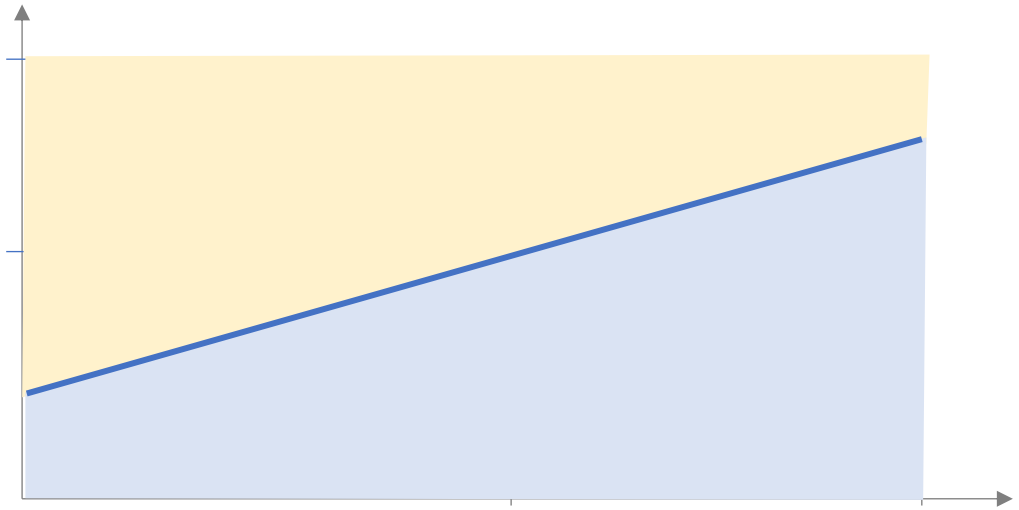
Stance detection: The task of determining whether someone is for or against a particular thing. We'll focus on "stance towards 4/574"



Adapted from: https://courses.cs.washington.edu/courses/cse416/21sp/

Sample task: stance detection

Stance detection: The task of determining whether someone is for or against a particular thing. We'll focus on "stance towards 4/574"

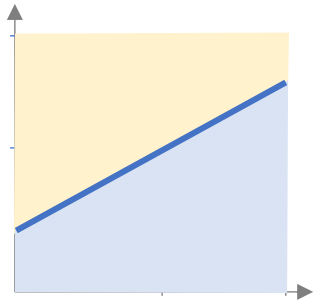


Adapted from: https://courses.cs.washington.edu/courses/cse416/21sp/

Sample task: stance detection

Stance detection: The task of determining whether someone is for or against a particular thing. We'll focus on "stance towards 4/574"

This class is garbage. The professor makes bad jokes and I can't read his handwriting



Arguably the greatest moments in human history have come when Kenny takes the floor for 4/574 each week

Adapted from: https://courses.cs.washington.edu/courses/cse416/21sp/

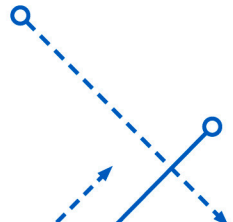
Stance detection in the real world

The prof's jokes are bad and he can't make a quiz without an error to save his life but I occasionally learn some stuff

Back to the example…

This class is garbage. The professor makes bad jokes and I can't read his handwriting

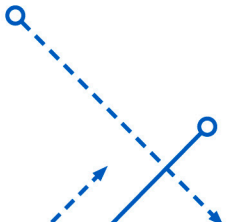
- How would you approach the task of stance detection? Specifically…
  - What would your **features** be?
  - How would you make decisions based on those features?
  - What **loss function** would you use?



Adapted from: https://courses.cs.washington.edu/courses/cse416/21sp/

Approach 1: Bag of words + Linear threshold classifier

1. Convert each course evaluation statement into a "bag of words" representation
2. Fix a weight for each word in terms of how having it in a sentence implies a positive/pro or negative/anti stance
3. Sum up the weights for all of the words to get a ***score***
4. If the ***score*** is > 0, predict "pro-5/474", otherwise, predict "anti"



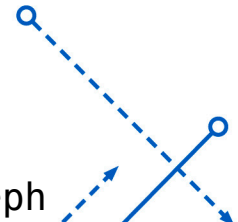
Adapted from: https://courses.cs.washington.edu/courses/cse416/21sp/

Approach 1, Step 1

This class is garbage. The professor makes bad jokes and I can't read his handwriting

The prof's jokes are bad and he can't make a quiz without an error to save his life but I occasionally learn some stuff

Arguably the greatest moments in human history have come when Kenny takes the floor for 4/574 each week



Approach 1, Step 2

- How might we get these values in the easiest possible way?
- … later… how we can learn them from data

Word	Weight (*w*)
garbage	-5.0
bad	-3.0
can't	-0.5
bad	-3.0
error	-2.1
learn	3.0
greatest	4.0
Arguably, human, professor, handwriting, …	0.0

Adapted from: https://courses.cs.washington.edu/courses/cse416/21sp/

Approach 1, Step 2

This class is garbage. The professor makes bad jokes and I can't read his handwriting

The prof's jokes are bad and he can't make a quiz without an error to save his life but I occasionally learn some stuff

Arguably the greatest moments in human history have come when Kenny takes the floor for 4/574 each week

Word	Weight (*w*)
garbage	-5.0
bad	-3.0
can't	-0.5
bad	-3.0
error	-2.1
learn	3.0
greatest	4.0
Arguably, human, professor, handwriting, …	0.0

Adapted from: https://courses.cs.washington.edu/courses/cse416/21sp/

Approach 1, Step 3

The prof's jokes are bad and he can't make a quiz without an error to save his life but I occasionally learn some stuff

Word	Weight (*w*)
garbage	-5.0
bad	-3.0
can't	-0.5
bad	-3.0
error	-2.1
learn	3.0
greatest	4.0
Arguably, human, professor, handwriting, …	0.0

Adapted from: https://courses.cs.washington.edu/courses/cse416/21sp/

Approach 1, Step 4

This class is garbage. The professor makes bad jokes and I can't read his handwriting

The prof's jokes are bad and he can't make a quiz without an error to save his life but I occasionally learn some stuff

Arguably the greatest moments in human history have come when Kenny takes the floor for 4/574 each week

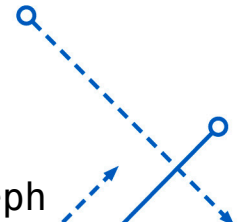
Adapted from: https://courses.cs.washington.edu/courses/cse416/21sp/

Cool!

- We have built our first classifier!
- **Quiz:** Did this classifier use (training) data at all?
- **How could it have used data to inform the model?**

…Put another way, how to learn word weights?

CIML, pg. 43



…Put another way, how to learn word weights?

An online algorithm to learn weights for the words…

The *perceptron* algorithm.

An early, well-known approach!

IMO, can complicate understanding at this point in the class

Algorithm 5 PERCEPTRONTRAIN(**D**, *MaxIter*)

```
w_d ← 0, for all d = 1 ... D                          // initialize weights
b ← 0                                                 // initialize bias
for iter = 1 ... MaxIter do
    for all (x,y) ∈ D do
        a ← Σ_{d=1}^{D} w_d x_d + b                   // compute activation for this example
        if ya ≤ 0 then
            w_d ← w_d + y x_d, for all d = 1 ... D    // update weights
            b ← b + y                                 // update bias
        end if
    end for
end for
return w_0, w_1, ..., w_D, b
```

Algorithm 6 PERCEPTRONTEST($w_0, w_1, \ldots, w_D, b, \hat{\mathbf{x}}$)

```
a ← Σ_{d=1}^{D} w_d x̂_d + b                           // compute activation for the test example
return SIGN(a)
```

CIML, pg. 43

Will any of this be relevant soon?

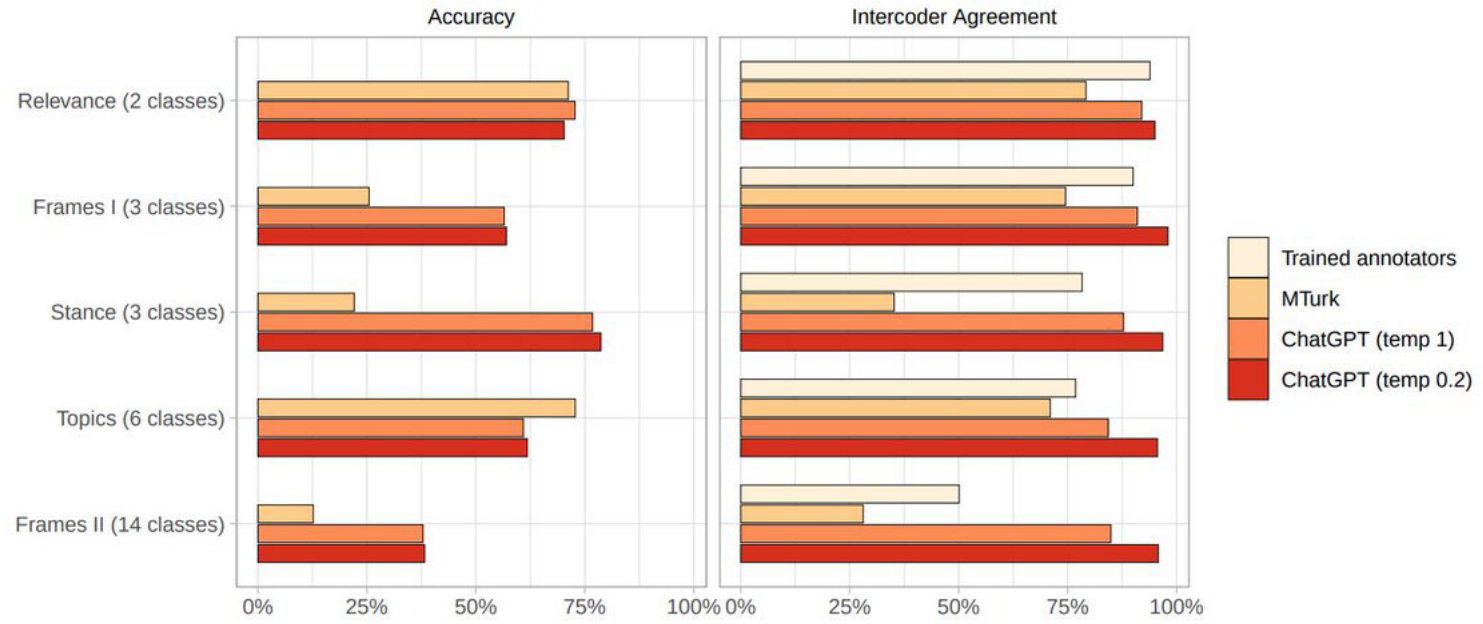


Figure 1: *ChatGPT zero-shot text annotation performance, compared to MTurk and trained annotators. ChatGPT's accuracy outperforms that of MTurk for four of the five tasks. ChatGPT's intercoder agreement outperforms that of both MTurk and trained annotators in all tasks. Accuracy means agreement with the trained annotators.*

OK!

What questions do you have?!