ILLINOIS INSTITUTE OF TECHNOLOGY

Sentiment analysis

CS-585

Natural Language Processing

Derrick Higgins

Sentiment analysis

- *Labeling the affective content of a text along a positive-negative scale; determining the speaker or author's orientation toward the topic of discussion*
- Research dates to ~2002: Bo Pang, Lillian Lee, & Shivakumar Vaithyanathan. *Thumbs Up? Sentiment Classification Using Machine Learning Techniques*
  Proceedings of EMNLP.
- AKA "opinion mining"

Example

This is the definitive movie version of Hamlet. Branagh cuts nothing, but there are no wasted moments. 👍

no comment - stupid movie, acting average or worse... screenplay - no sense at all... SKIP IT! 👎

This movie is terrible but it has some good effects. 👎

I don't know why I like this movie so well, but I never get tired of watching it. 👍

Brilliant movie. The drawings were just amazing. Too bad it ended before it begun. I´ve waited 21 years for a sequel, but nooooo!!! 👍

I wouldn't rent this one even on dollar rental night. 👎

Applications

- "Voice of the customer"
  - Are my customers happy?
  - Are my products and services well-regarded?
- Customer service
  - Are my representatives dealing with my customers in a respectful and helpful way?
- E-commerce
  - How satisfied are customers with products or categories of products?

Sample tasks

- Review classification
  - Movies, TV shows, podcasts, …
  - Restaurants, bars
  - Products sold online
- Social media monitoring
- Customer/representative interactions
  - SMS chats
  - Call transcripts
  - Email
  - Snail mail

Predict star rating

Predict manual annotations

Sentiment and *subjectivity*

- Sentiment is *affective polarity* – positivity or negativity of the content
- But what if it's just factual, not evaluative?
  - "An oxygen atom has 8 protons."
  - "I had pizza for lunch."
- In a sentiment task, we could say that these sentences are neutral. But sometimes we also make a binary distinction between subjective and non-subjective texts. Only subjective texts can have sentiment.

Sentiment and *emotion*

- Emotion is clearly related to sentiment, but more fine-grained
- Different classification schemes
  - Ekman (1992): anger, disgust, fear, happiness, sadness, surprise
  - Others include even more distinctions
- May be purely text-based or *multimodal* (incorporating visual and auditory evidence)

Sentiment and *stance*

- *Stance*: the side of an issue taken by a speaker/writer
- Usually binary; not necessarily aligned with sentiment
- E.g.,
  - Interventionist vs. isolationist
  - Supporter / opponent of action against climate change
  - Pro or anti-GMO
  - iPhone vs. Android

Sentiment and *aspect*

- Sentiment is about the affective orientation of a text
- But the polarity can depend on the specific subject of the evaluation

The culinary experience was as excellent as other reviewers have noted, but we practically had to send a search party to locate our waiter when needed.

While an all-star cast made the most of the material, ultimately they couldn't make up for the inane script they were given.

Food: positive
Service: negative

Acting: positive
Writing: negative

MODELS FOR SENTIMENT ANALYSIS

Sentiment analysis as text categorization

- Is sentiment a classification task or a regression task?
  - Are texts either negative or positive, or can we put them on a scale?

The chef's culinary mastery transported me to a hitherto unexplored level of epicurean euphoria.

I enjoyed my meal.

I visited a restaurant on my trip to London.

The food was not great.

It was the most stereotypically British concoction of dreadful and unidentifiable boiled ingredients you could imagine.

Sentiment analysis as text categorization

- Sentiment is usually treated as a classification task, even though intuitively there are gradations
- How many classes?
  - Positive / negative (particularly in combination with subjectivity)
  - Positive / negative / neutral
  - Mixed / indeterminate?

Lexicon-based sentiment analysis

- Bing Liu's sentiment lexicon (one of a few)

a+	work	2-faced	wretchedness
abound	workable	2-faces	wrinkle
abounds	worked	abnormal	wrinkled
abundance	works	abolish	wrinkles
abundant	world-famous	abominable	wrip
accessable	worth	abominably	wripped
accessible	worth-while	abominate	wripping
acclaim	worthiness	abomination	writhe
acclaimed	worthwhile	abort	wrong
acclamation	worthy	aborted	wrongful
accolade	wow	aborts	wrongly
accolades	wowed	abrade	wrought
accommodative	wowing	abrasive	yawn
accomodative	wows	abrupt	zap
accomplish	yay	abruptly	zapped
accomplished	youthful	abscond	zaps
accomplishment	zeal	absence	zealot
accomplishments	zenith	absent-minded	zealous
accurate	zest	absentee	zealously
accurately	zippy	absurd	zombie

https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

Lexicon-based sentiment analysis

- To use a sentiment lexicon
  - Just count the positive/negative words (or sum the scores, if your lexicon has them)
  - OR do normal text classification, and use the lexicon as your vocabulary
- Pros
  - Curated list makes it less likely that the model will learn "weird" correlations (non-emotional terms as predictors of sentiment)
- Cons
  - Subjective decisions in list creation
  - Limits sophistication of sentiment classifier

Bag-of-words sentiment modeling

- Alternative (or complement) to lexicon-based sentiment analysis
- Given some labeled training data, just build a text categorization model (naïve Bayes, logistic regression, etc.) to predict the correct label (positive or negative) for a text
- Pros
  - Resulting classifier will be well-suited to domain of interest
  - Can leverage state-of-the-art machine learning for high accuracy
- Cons
  - Some of the relationships learned by the model may be questionable from a validity perspective (see notebook)
  - Finding good training data can be a challenge

Annotation for sentiment analysis

- Where to get training data for sentiment?
  1. Star ratings / likes / net promoter score
  2. Manual annotation (human judgements collected specifically for model training)

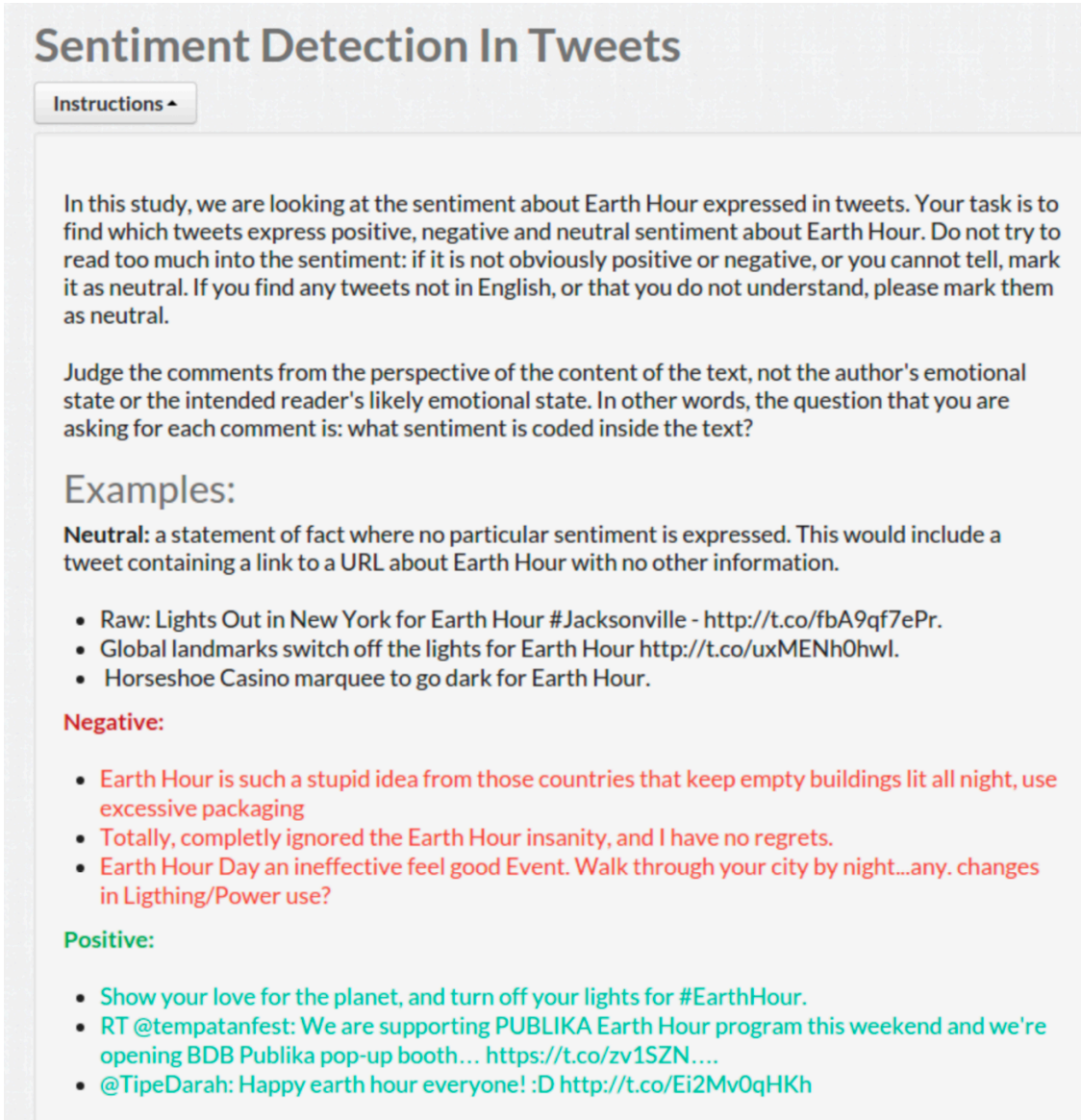
- Unfortunately, the contexts in which we have a lot of training data are also the contexts in which sentiment models add the least value
  - If your reviews already have star ratings, who needs sentiment?

Annotation for sentiment analysis

- Collecting annotations for modeling is a lot of work!
- At least a thousand or so texts must be labeled to have a big enough train/test set
- Some have to be labeled by multiple annotators to ensure the labeling is reliable
- Substantial work must be done up front
  - To create annotation guidelines explaining the task
  - To create an interface in which annotators can record their judgements

Annotation for sentiment analysis

- For example…
- Examples with important edge cases to ensure consistency
- Monitoring of annotator agreement



Sentiment Detection In Tweets

Instructions

In this study, we are looking at the sentiment about Earth Hour expressed in tweets. Your task is to find which tweets express positive, negative and neutral sentiment about Earth Hour. Do not try to read too much into the sentiment: if it is not obviously positive or negative, or you cannot tell, mark it as neutral. If you find any tweets not in English, or that you do not understand, please mark them as neutral.

Judge the comments from the perspective of the content of the text, not the author's emotional state or the intended reader's likely emotional state. In other words, the question that you are asking for each comment is: what sentiment is coded inside the text?

Examples:

Neutral: a statement of fact where no particular sentiment is expressed. This would include a tweet containing a link to a URL about Earth Hour with no other information.

- Raw: Lights Out in New York for Earth Hour #Jacksonville - http://t.co/fbA9qf7ePr.
- Global landmarks switch off the lights for Earth Hour http://t.co/uxMENh0hwI.
- Horseshoe Casino marquee to go dark for Earth Hour.

Negative:

- Earth Hour is such a stupid idea from those countries that keep empty buildings lit all night, use excessive packaging
- Totally, completly ignored the Earth Hour insanity, and I have no regrets.
- Earth Hour Day an ineffective feel good Event. Walk through your city by night...any. changes in Ligthing/Power use?

Positive:

- Show your love for the planet, and turn off your lights for #EarthHour.
- RT @tempatanfest: We are supporting PUBLIKA Earth Hour program this weekend and we're opening BDB Publika pop-up booth… https://t.co/zv1SZN….
- @TipeDarah: Happy earth hour everyone! :D http://t.co/Ei2Mv0qHKh

Maynard, Diana & K. Bontcheva. 2016. *Challenges of Evaluating Sentiment Analysis Tools on Social Media*. Proceedings of LREC.

Annotation for sentiment analysis

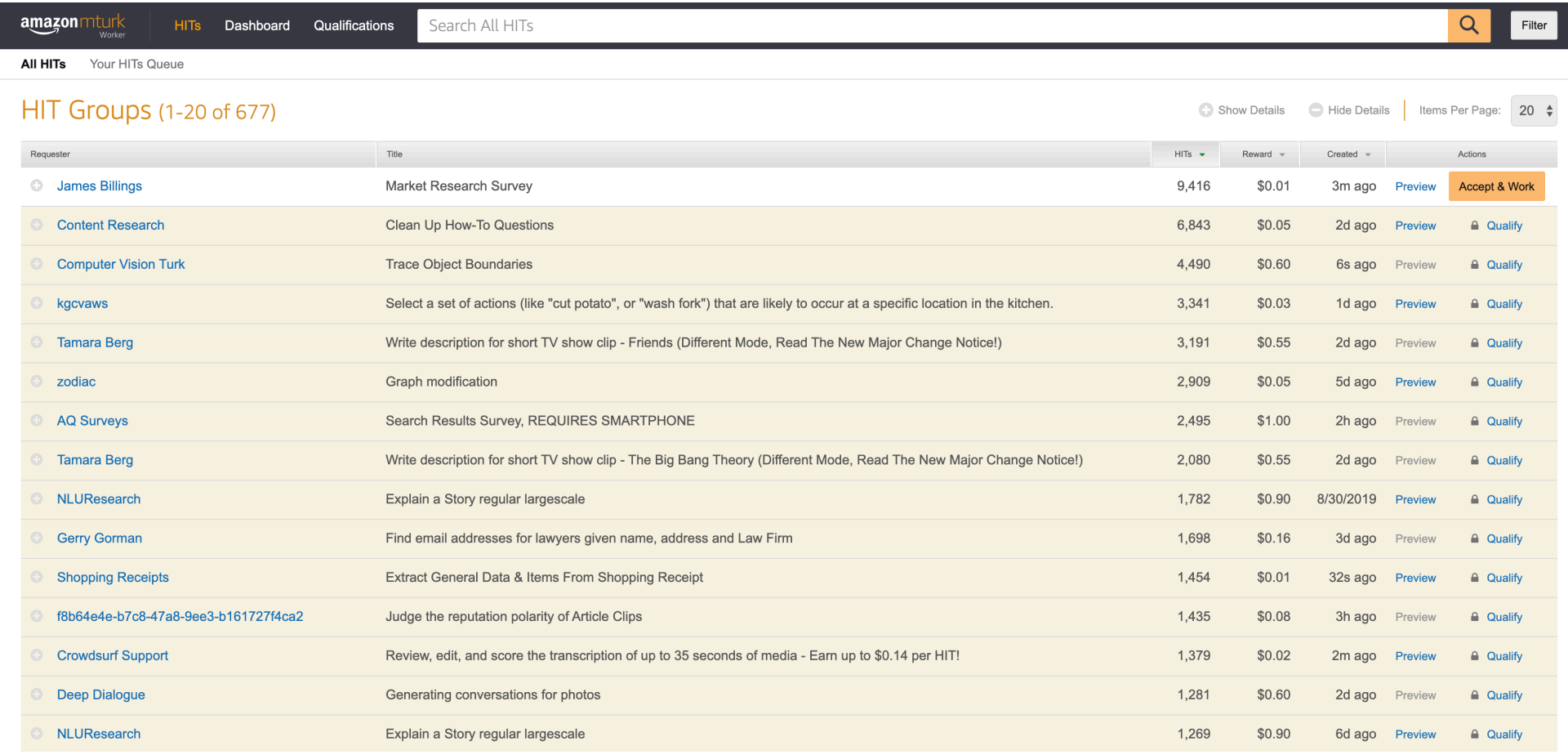
Who does the annotation work?

A. Scientists/students
   - High quality
   - NOT scalable
   - Free

B. Crowdsourced workers (e.g., Amazon Mechanical Turk)
   - Low quality
   - Scalable
   - Inexpensive

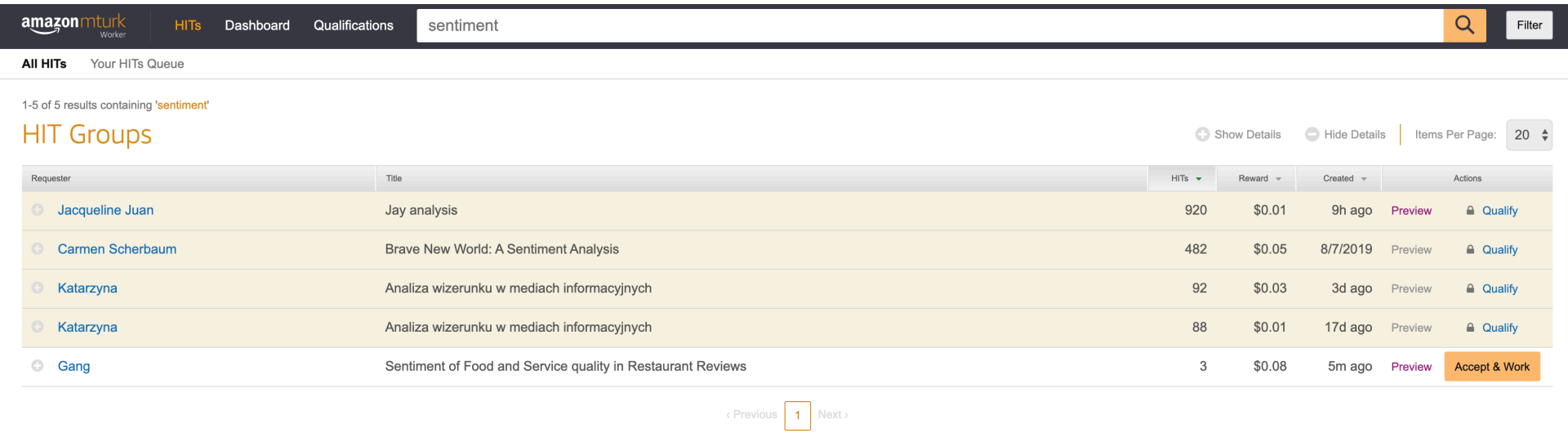
C. Dedicated workers
   - High quality
   - Scalable
   - Expensive

Annotation for sentiment analysis

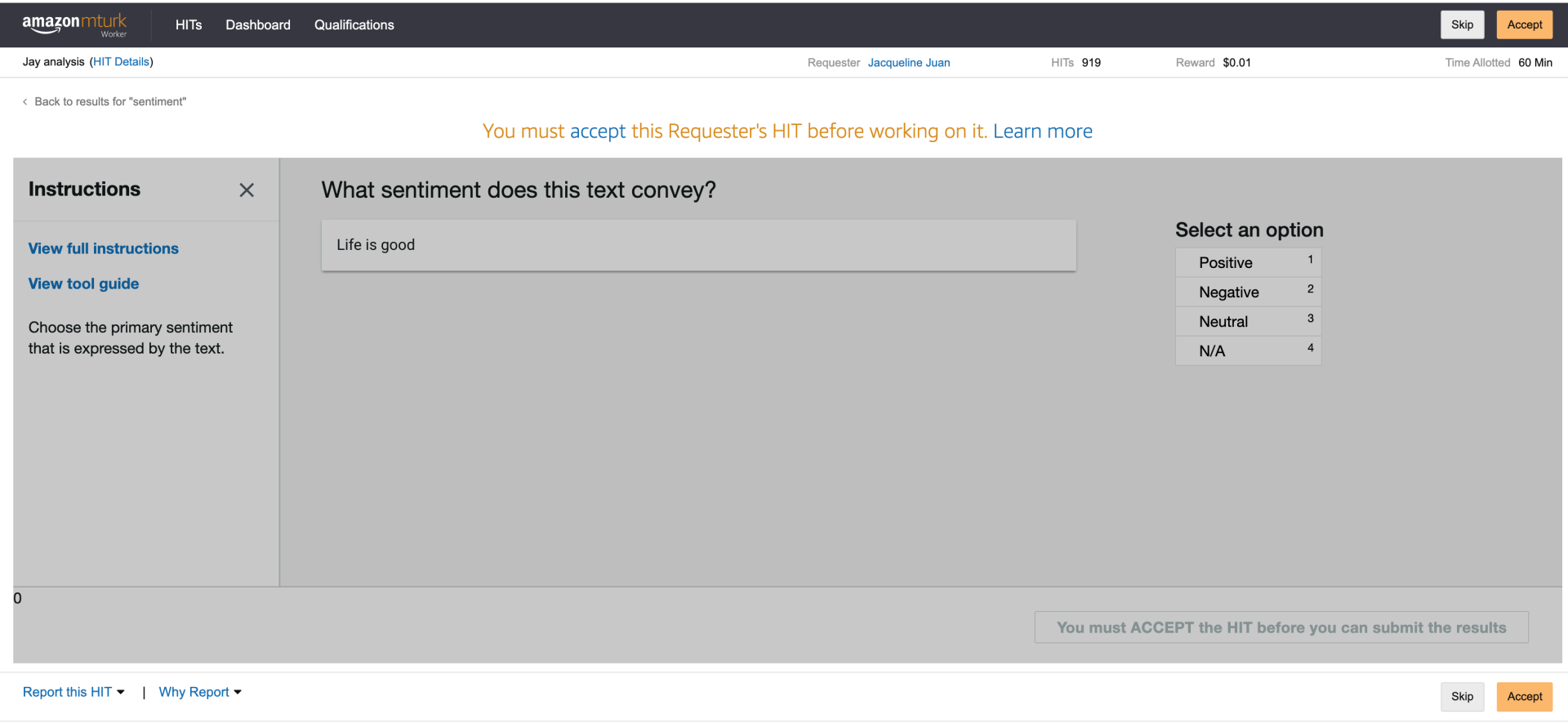


Requester	Title	HITs	Reward	Created	Actions
James Billings	Market Research Survey	9,416	$0.01	3m ago	Preview, Accept & Work
Content Research	Clean Up How-To Questions	6,843	$0.05	2d ago	Preview, Qualify
Computer Vision Turk	Trace Object Boundaries	4,490	$0.60	6s ago	Preview, Qualify
kgcvaws	Select a set of actions (like "cut potato", or "wash fork") that are likely to occur at a specific location in the kitchen.	3,341	$0.03	1d ago	Preview, Qualify
Tamara Berg	Write description for short TV show clip - Friends (Different Mode, Read The New Major Change Notice!)	3,191	$0.55	2d ago	Preview, Qualify
zodiac	Graph modification	2,909	$0.05	5d ago	Preview, Qualify
AQ Surveys	Search Results Survey, REQUIRES SMARTPHONE	2,495	$1.00	2h ago	Preview, Qualify
Tamara Berg	Write description for short TV show clip - The Big Bang Theory (Different Mode, Read The New Major Change Notice!)	2,080	$0.55	2d ago	Preview, Qualify
NLUResearch	Explain a Story regular largescale	1,782	$0.90	8/30/2019	Preview, Qualify
Gerry Gorman	Find email addresses for lawyers given name, address and Law Firm	1,698	$0.16	3d ago	Preview, Qualify
Shopping Receipts	Extract General Data & Items From Shopping Receipt	1,454	$0.01	32s ago	Preview, Qualify
f8b64e4e-b7c8-47a8-9ee3-b161727f4ca2	Judge the reputation polarity of Article Clips	1,435	$0.08	3h ago	Preview, Qualify
Crowdsurf Support	Review, edit, and score the transcription of up to 35 seconds of media - Earn up to $0.14 per HIT!	1,379	$0.02	2m ago	Preview, Qualify
Deep Dialogue	Generating conversations for photos	1,281	$0.60	2d ago	Preview, Qualify
NLUResearch	Explain a Story regular largescale	1,269	$0.90	6d ago	Preview, Qualify

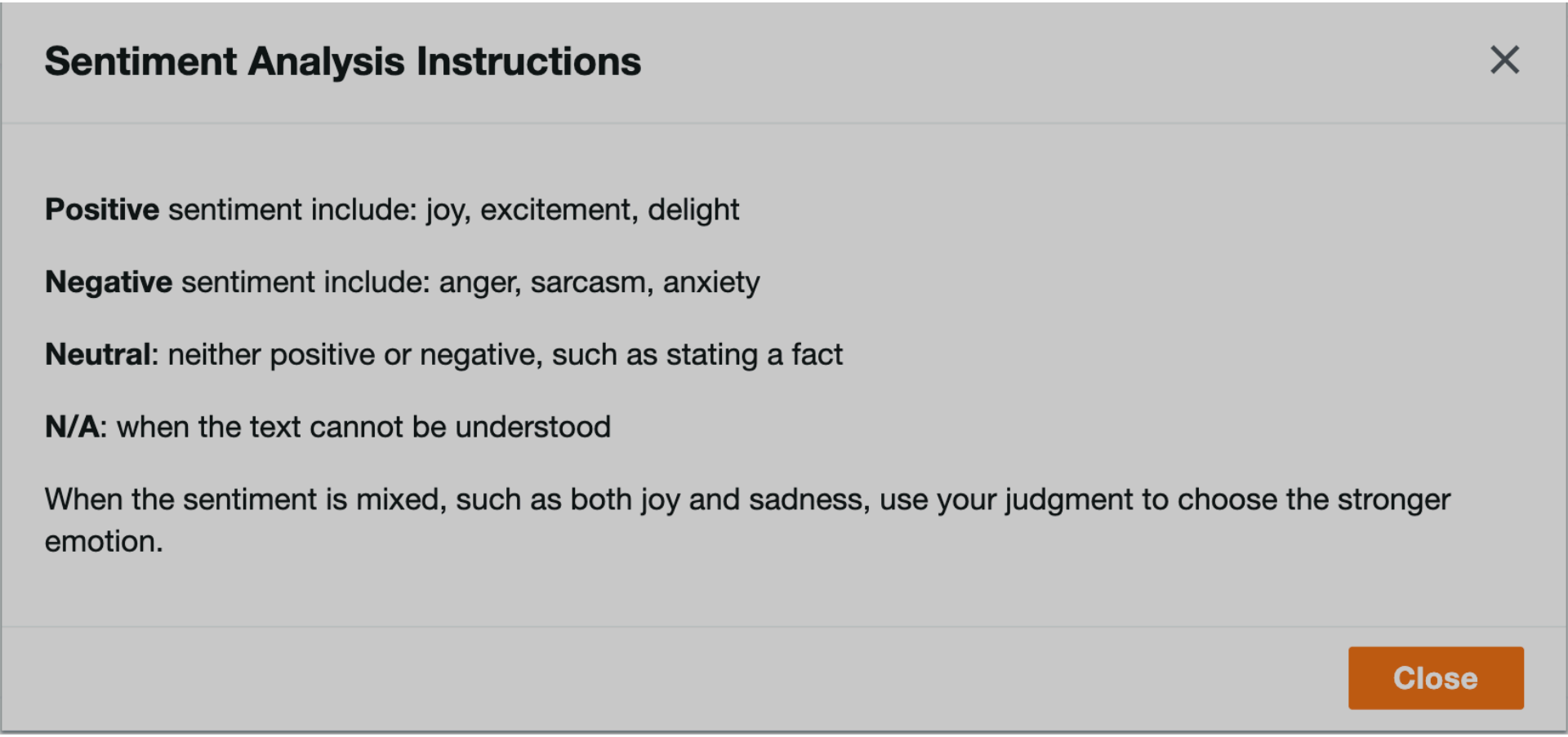
Annotation for sentiment analysis



Annotation for sentiment analysis



Annotation for sentiment analysis



SENTIMENT AND ASPECT

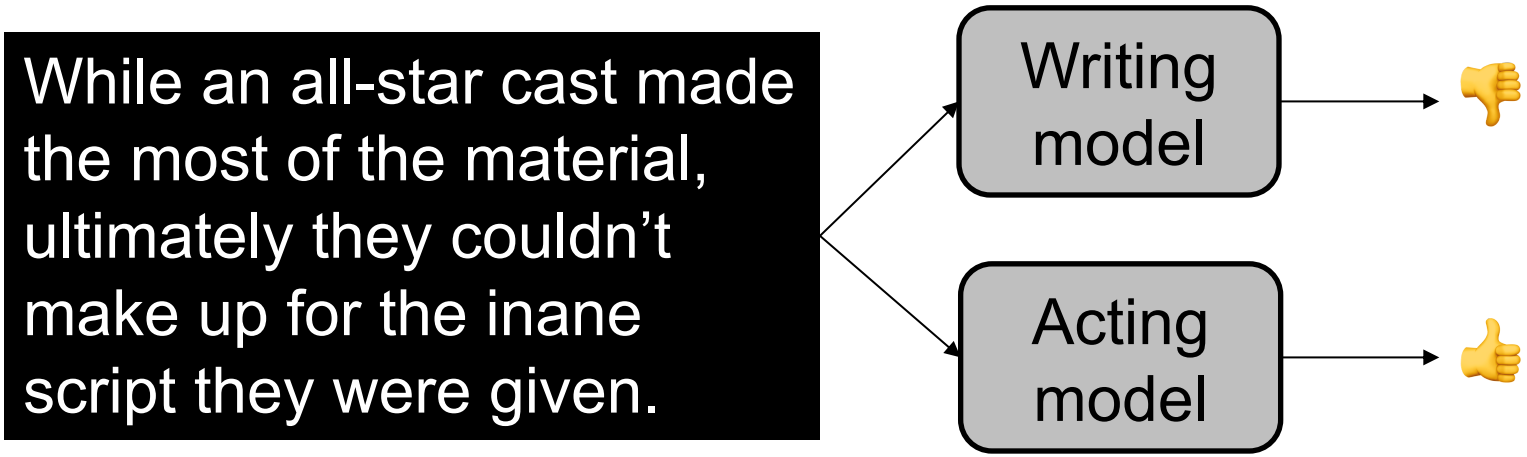
Sentiment and *aspect*

- *Aspect* is the characteristic or object of opinion focus within a text
- Aspect is typically domain-specific

Domain	Sample aspects
Films and TV	Screenplay, acting, special effects
Phones	Screen, battery, apps, price
Restaurants	Food, service, atmosphere, price
Healthcare	Bedside manner, treatment, scheduling
Banking	Fees, services, interest rates, perks

Sentiment and *aspect*

- If aspect is simply a category of opinion, it can be handled with N separate classifiers for N categories



Sentiment and *aspect*

- Aspect-based sentiment may also be relative to a specific target phrase, e.g:

One Iowa voter said she was excited to see Kamala Harris in person, but was still on the fence about who to support in the caucus.

The Bears are a heavy underdog against the Super Bowl champion Patriots on Sunday.

GOOG rebounded in afternoon trading, with tech stocks rallying on news of the agreement's approval.

No longer just a bag-of-words problem

CHALLENGES FOR SENTIMENT ANALYSIS

Domain specificity

- Off-the-shelf sentiment modeling is widely sold, and many business users assume it is sufficient for their needs
- But the definition of *sentiment* can be task-specific

Task	Sentiment
Product review rating	Would the reviewer buy again?
Social media monitoring	Is my brand getting positive mentions?
Customer service call monitoring	Was the customer's problem resolved?
Customer service call monitoring	Was the service agent polite and respectful?
Financial NLP	Are markets bullish on this company?

Domain specificity

- …and the types of words and phrases that count as negative are highly domain-dependent

Domain	Negative terms
Films and TV	Melodrama, sequel, rehash
E-commerce	Return, porch pirate, 404
Restaurants	Noisy, insects, wait, grease
Insurance	Denial, copayment, preauthorization
Construction	Estimate, overrun, subcontractor

Need for interpretability

- Surprising words sometimes show up with sentiment polarity – they tend to correlate with negative discussion (for example), but aren't themselves negative

earthquake hospital hurricane	ex-wife politician salesperson	vacation puppy popsicle

- This can be problematic if we want to use sentiment scores to attribute opinions to people – the evidence will not stand up to review

Aggregation of sentiment scores

- Many sentiment applications involve tracking across larger groups of people
  - Are my product's ratings improving or declining?
  - Are my customers more satisfied than they were this time last year?
  - Are my rural customers more or less satisfied than my urban customers?
- It's not always clear how to get from a text-level sentiment score to the metric you really want
  - Average sentiment per sentence for all customer service calls doesn't tell you if the customer got their problem resolved
  - Similarly, a text could have a lot of negative content, but still end with a positive customer experience

Is a bag of words enough?

- We've been treating sentiment analysis as a standard text categorization task
- …and a bag-of-words representation is common for text categorization.
- But is it really appropriate for sentiment?

I would never buy this product again. It clearly failed under high-stress testing in my home.

I would clearly buy this product again. It never failed under high-stress testing in my home.

Is a bag of words enough?

- Sentiment can change markedly based on seemingly small differences between texts
  - Presence and focus of negation
  - Counterfactual/conditional contexts
  - Swapping of argument noun phrases

- More than other tasks, sentiment analysis might benefit from information about word order

Beyond bag of words for sentiment

- Neural models are a natural fit
- Socher et al. (2013)
  - Phrase-level sentiment scores for over 215K phrases (≈12K sentences)
  - Recursive architecture predicts sentiment for each constituent of a syntactic structure, until tree root (full sentence) is reached
  - Detailed analysis of how linguistic cues to sentiment are captured by the model
  - Full-featured demo, code, and corpus at the project site: https://nlp.stanford.edu/sentiment/

Socher, Richard, et al. 2013. *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank*. In EMNLP.

Beyond bag of words for sentiment

Model	Fine-grained		Positive/Negative	
	All	Root	All	Root
NB	67.2	41.0	82.6	81.8
SVM	64.3	40.7	84.6	79.4
BiNB	71.0	41.9	82.7	83.1
VecAvg	73.3	32.7	85.1	80.1
RNN	79.0	43.2	86.1	82.4
MV-RNN	78.7	44.4	86.8	82.9
RNTN	**80.7**	**45.7**	**87.6**	**85.4**

Table 1: Accuracy for fine grained (5-class) and binary predictions at the sentence level (root) and for all nodes.

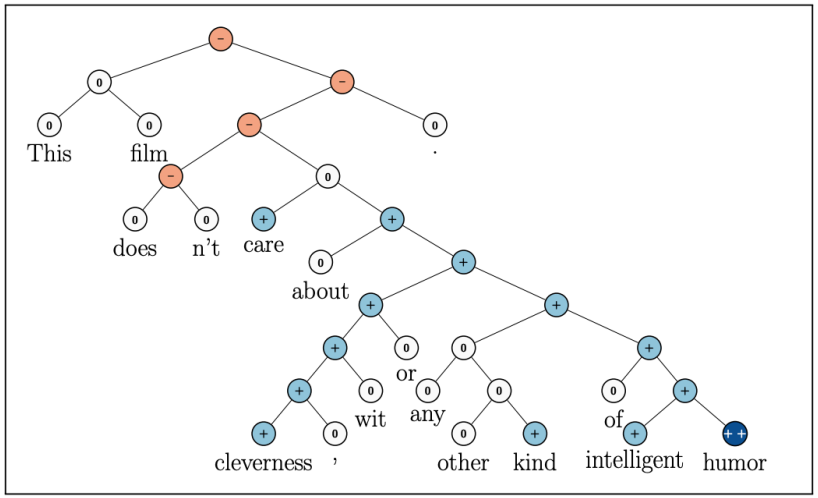


Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (– –, –, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence.

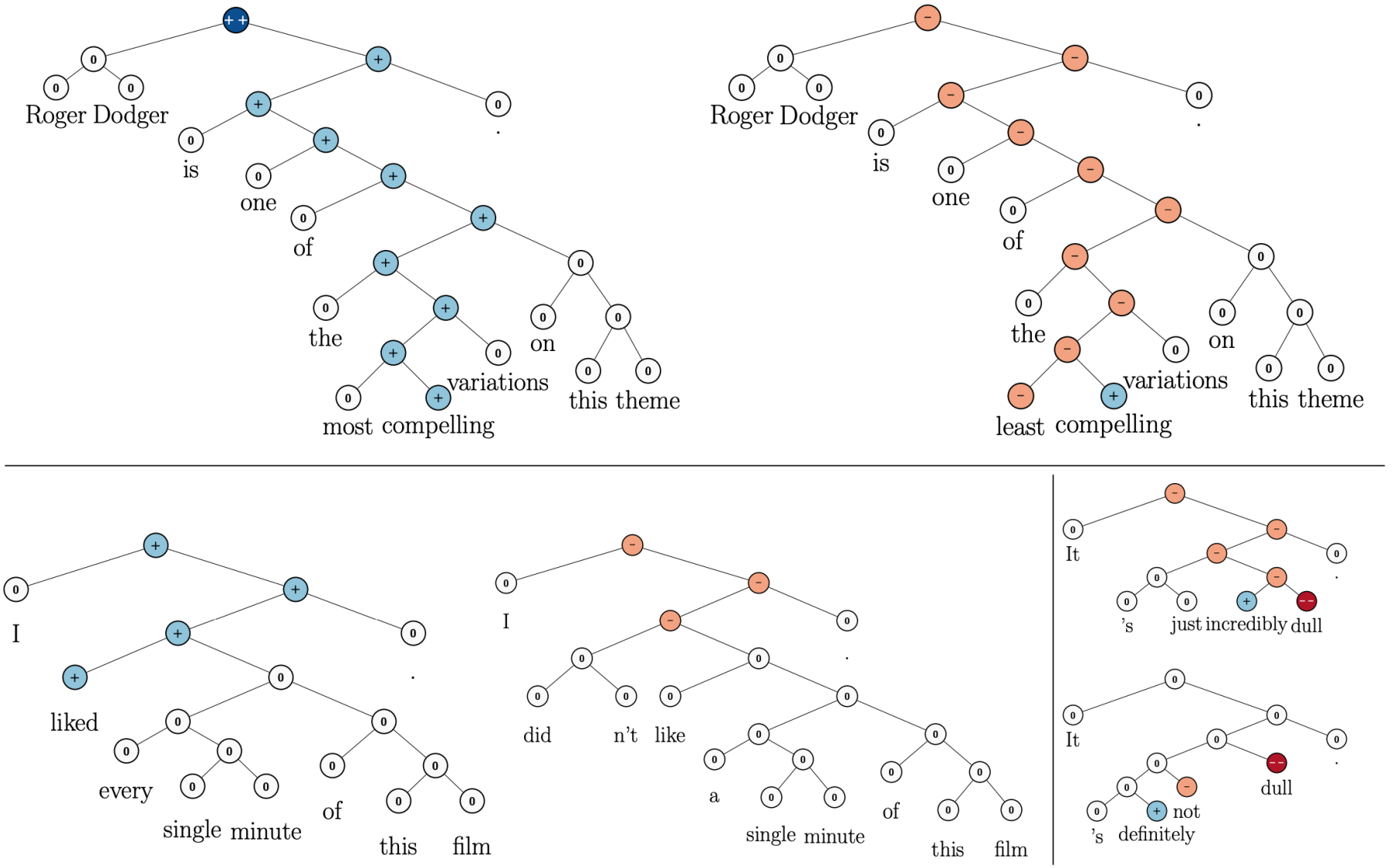


Figure 9: RNTN prediction of positive and negative (bottom right) sentences and their negation.

Sentiment demo

[Notebook]