CS378: Natural Language Processing

Lecture 9: Meta NLP

Eunsol Choi

Some slides adapted from Yoav Artzi / Greg Durrett

Logistics

- HW2 due in a week!

Expectation Maximization

- Learning objective:

$$L(\theta) = \sum_i \log \sum_{y \in \mathcal{Y}} P(x_i, y \mid \theta)$$

- The EM (Expectation Maximization) algorithm is a method for finding

$$\theta_{MLE} = \arg\max_\theta L(\theta) = \arg\max_\theta \sum_i \log \sum_{y \in \mathcal{Y}} P(x_i, y \mid \theta)$$

- We will look into EM in HMM!

$$p(x_1 \dots x_n, y_1 \dots y_n) = q(STOP|y_n) \prod_{i=1}^{n} q(y_i|y_{i-1}) e(x_i|y_i)$$

EM Intuition

- What we want is…

$$p(y_i|x_1...x_n) = \frac{p(x_1...x_n, y_i)}{p(x_1...x_n)}$$

- We can compute:

$$\text{(expected) count(NN)} = \sum_i p(y_i = \text{NN}|x_1...x_n)$$

- If we have….

$$p(y_i y_{i+1}|x_1...x_n) = \frac{p(x_1...x_n, y_i, y_{i+1})}{p(x_1...x_n)}$$

- Then we can compute expected transition counts:

$$\text{(expected) count(NN} \rightarrow \text{VB)} = \sum_i p(y_i = \text{NN}, y_{i+1} = \text{VB}|x_1...x_n)$$

- Above marginals can be computed from followings:

$$p(x_1...x_n, y_i) = \alpha(i, y_i)\beta(i, y_i)$$

$$p(x_1...x_n, y_i, y_{i+1}) = \alpha(i, y_i)q(y_{i+1}|y_i)e(x_{i+1}|y_{i+1})\beta(i+1, y_{i+1})$$

Toy Example: Ice Cream Climatology

- You are studying global warming. You cannot find records of weather, but you can find records of how much ice cream was consumed each day. Can you estimate the weather history from the ice cream history?
  - Observations (x): Number of ice cream purchase
    - {1, 2, 3}
  - State (y): Weather
    - {C (cold), H (hot)}

Toy Example: Ice Cream Climatology

If today is cold (C) or hot (H), how many cones did I prob. eat?

	P(..|C)	P(..|H)
P(1|..)	0.7	0.1
P(2|..)	0.2	0.2
P(3|..)	0.1	0.7

	P(..|C)	P(..|H)	P(..|start)
P(C|..)	0.8	0.1	0.5
P(H|..)	0.1	0.8	0.5
P(Stop|..)	0.1	0.1	0

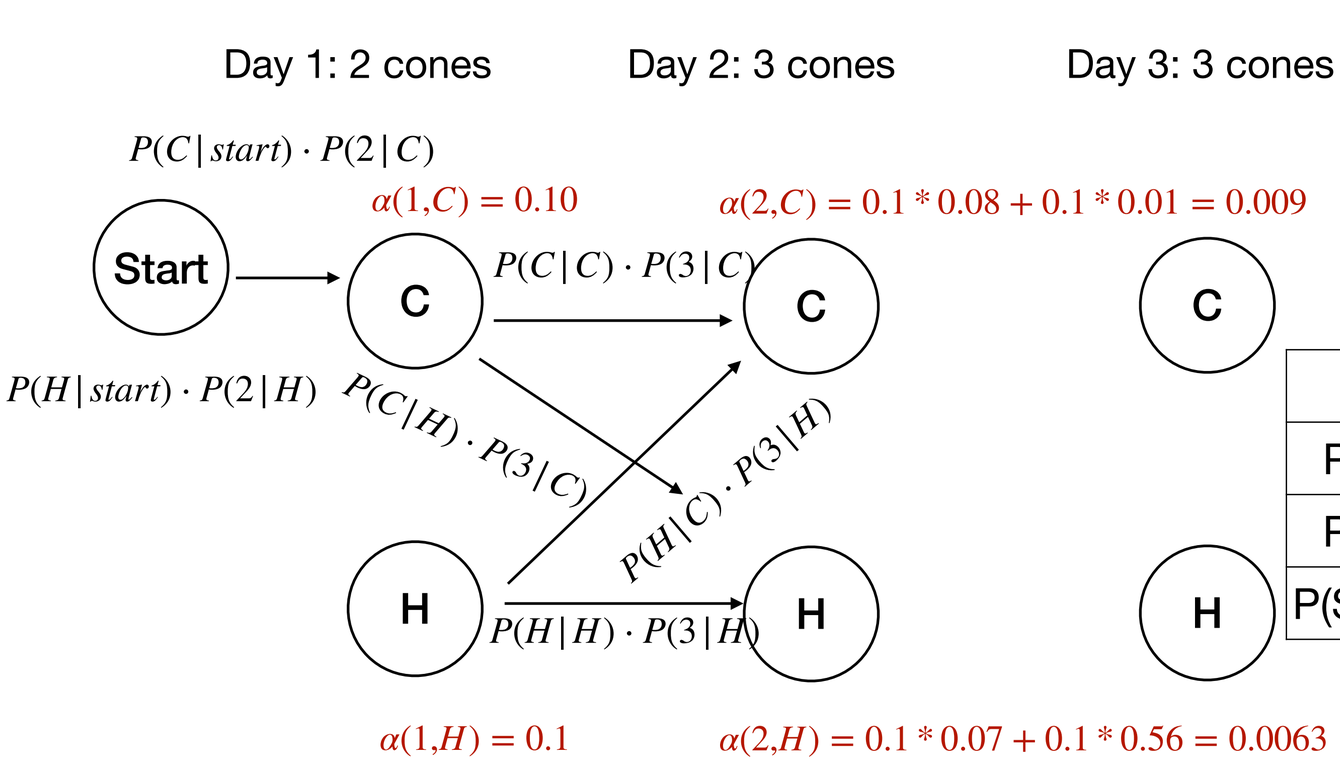
If today is cold (C) or hot (H),what will tomorrow's weather be?

- Maximum Likelihood Parameters (supervised):

$$e_{ML}(x|y) = \frac{c(y,x)}{c(y)} \qquad q_{ML}(y_i|y_{i-1}) = \frac{c(y_{i-1},y_i)}{c(y_{i-1})}$$

- Now, we do not have the weather record (real counts), so we start with guessed probability table, and compute **expected counts**

Toy Example: Ice Cream Climatology



	P(..|C)	P(..|H)
P(1|..)	0.7	0.1
P(2|..)	0.2	0.2
P(3|..)	0.1	0.7

	P(..|C)	P(..|H)	P(..|start)
P(C|..)	0.8	0.1	0.5
P(H|..)	0.1	0.8	0.5
P(Stop|..)	0.1	0.1	0

Day #	Ice Creams	p(□C)	p(□H)
1	2	0.129	0.871
2	3	0.023	0.977
3	3	0.011	0.989
4	2	0.027	0.973
5	3	0.013	0.987
6	2	0.032	0.968
7	3	0.022	0.978
8	2	0.069	0.931
9	2	0.089	0.911
10	3	0.082	0.918
11	1	0.248	0.752
12	3	0.144	0.856
13	3	0.221	0.779
14	1	0.887	0.113
15	1	0.98	0.02
16	1	0.991	0.009
17	2	0.977	0.023
18	1	0.994	0.006
19	1	0.994	0.006
20	1	0.977	0.023
21	3	0.857	0.143
22	1	0.962	0.038
23	2	0.961	0.039
24	1	0.989	0.011
25	1	0.985	0.015
26	1	0.926	0.074
27	2	0.507	0.493
28	3	0.087	0.913
29	3	0.032	0.968
30	2	0.053	0.947
31	3	0.045	0.955
32	2	0.146	0.854
33	2	0.225	0.775
		14.679	**18.321**

- We compute forward and backward, from which we compute P(y_i|**x**) at each day (i).

$$p(x_1 \ldots x_n, y_i) = \alpha(i, y_i)\beta(i, y_i)$$

Toy Example: Ice Cream Climatology

Day #	Ice Creams	p(□C)	p(□H)	p(□C,1)	p(□C,2)	p(□C,3)	p(□H,1)	p(□H,2)	p(□H,3)	p(C□C)	p(H□C)	p(C□H)	p(H□H)
1	2	0.129	0.871	0	0.129	0	0	0.871	0	#N/A	#N/A	#N/A	#N/A
2	3	0.023	0.977	0	0	0.023	0	0	0.977	0.021	0.003	0.109	0.868
3	3	0.011	0.989	0	0	0.011	0	0	0.989	0.006	0.005	0.017	0.972
4	2	0.027	0.973	0	0.027	0	0	0.973	0	0.006	0.021	0.005	0.969
5	3	0.013	0.987	0	0	0.013	0	0	0.987	0.007	0.005	0.02	0.968
6	2	0.032	0.968	0	0.032	0	0	0.968	0	0.008	0.024	0.005	0.963
7	3	0.022	0.978	0	0	0.022	0	0	0.978	0.012	0.009	0.02	0.959
8	2	0.069	0.931	0	0.069	0	0	0.931	0	0.017	0.052	0.005	0.927
9	2	0.089	0.911	0	0.089	0	0	0.911	0	0.05	0.038	0.018	0.893
10	3	0.082	0.918	0	0	0.082	0	0	0.918	0.057	0.025	0.031	0.886
11	1	0.248	0.752	0.248	0	0	0.752	0	0	0.077	0.171	0.005	0.747
12	3	0.144	0.856	0	0	0.144	0	0	0.856	0.131	0.013	0.117	0.739
13	3	0.221	0.779	0	0	0.221	0	0	0.779	0.128	0.093	0.016	0.762
14	1	0.887	0.113	0.887	0	0	0.113	0	0	0.221	0.666	0.001	0.113
15	1	0.98	0.02	0.98	0	0	0.02	0	0	0.884	0.095	0.003	0.018
16	1	0.991	0.009	0.991	0	0	0.009	0	0	0.975	0.016	0.005	0.005
17	2	0.977	0.023	0	0.977	0	0	0.023	0	0.973	0.004	0.018	0.005
18	1	0.994	0.006	0.994	0	0	0.006	0	0	0.974	0.019	0.003	0.004
19	1	0.994	0.006	0.994	0	0	0.006	0	0	0.989	0.005	0.005	0.002
20	1	0.977	0.023	0.977	0	0	0.023	0	0	0.974	0.003	0.019	0.004
21	3	0.857	0.143	0	0	0.857	0	0	0.143	0.855	0.002	0.122	0.021
22	1	0.962	0.038	0.962	0	0	0.038	0	0	0.853	0.109	0.004	0.034
23	2	0.961	0.039	0	0.961	0	0	0.039	0	0.944	0.017	0.018	0.021
24	1	0.989	0.011	0.989	0	0	0.011	0	0	0.957	0.032	0.004	0.008
25	1	0.985	0.015	0.985	0	0	0.015	0	0	0.978	0.007	0.011	0.005
26	1	0.926	0.074	0.926	0	0	0.074	0	0	0.924	0.003	0.061	0.012
27	2	0.507	0.493	0	0.507	0	0	0.493	0	0.505	0.001	0.421	0.072
28	3	0.087	0.913	0	0	0.087	0	0	0.913	0.085	0.002	0.421	0.492
29	3	0.032	0.968	0	0	0.032	0	0	0.968	0.026	0.006	0.061	0.907
30	2	0.053	0.947	0	0.053	0	0	0.947	0	0.022	0.031	0.01	0.937
31	3	0.045	0.955	0	0	0.045	0	0	0.955	0.028	0.017	0.025	0.931
32	2	0.146	0.854	0	0.146	0	0	0.854	0	0.04	0.106	0.005	0.849
33	2	0.225	0.775	0	0.225	0	0	0.775	0	0.13	0.095	0.016	0.759
		14.679	**18.321**	**9.931**	**3.212**	**1.537**	**1.069**	**7.788**	**9.463**	**12.855**	**1.695**	**1.599**	**15.85**

- With the expected counts for hot and cold days for each day, we compute the following

$$p(x_1 ... x_n, y_i) = \alpha(i, y_i)\beta(i, y_i)$$

$$p(x_1 ... x_n, y_i, y_{i+1}) = \alpha(i, y_i)q(y_{i+1}|y_i)e(x_{i+1}|y_{i+1})\beta(i+1, y_{i+1})$$

- Use these values that count to re-comput transition probability and emission probability

	P(..|C)	P(..|H)
P(1|..)	0.6765	0.0584
P(2|..)	0.2188	0.4251
P(3|..)	0.1047	0.5165

	P(..|C)	P(..|H)	P(..|start)
P(C|..)	0.8757	0.0925	0.1291
P(H|..)	0.109	0.8652	0.8709
P(Stop|..)	0.0153	0.0423	0

Quiz: p(S1) vs. p(S2)

- S1 = Colorless green ideas sleep furiously.
- S2 = Furiously sleep ideas green colorless
  - “It is fair to assume that neither sentence (S1) nor (S2) had ever occurred in an English discourse. Hence, in any statistical model for grammaticalness, these sentences will be ruled out on identical grounds as equally "remote" from English” (Chomsky 1957)

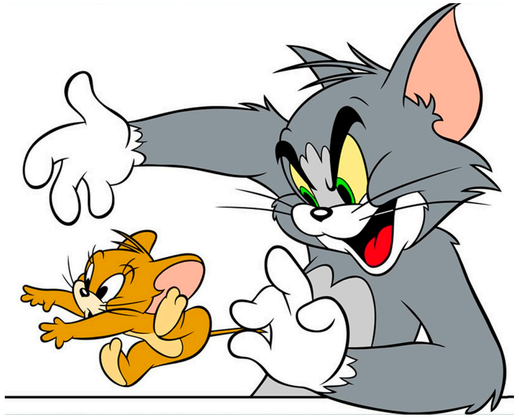
- How would p(S1) and p(S2) compare based on (smoothed) bigram language models?
- How would p(S1) and p(S2) compare based on marginal probability based on POS-tagging HMMs?
  - i.e., marginalized over all possible sequences of POS tags

Next Two Lectures

- Meta-NLP
- Getting started with NLP research project
  - Where should we start?
  - Dataset
  - Evaluation
  - Model
- Ethics in NLP

NLP as a Arms-Race

Data



Model

[Marcus et al 1994, Koehn and Monz 2006, Pang and Lee 2008, Rajpurkar et al 2016]

[Williams 1992, Lafferty et al 2001, Collins 2002, Roth and Yih 2005, Vinyals et al 2015]

11

NLP as a Arms-Race

Data

- Benchmark evaluation dataset
- Paired with training dataset
- Examples:
  - Example Translation Pairs (WMT)
  - Parse Trees for Language (PennTreebank)
  - IMDB Movie Review Dataset
  - GeoQuery Dataset

Model

[Marcus et al 1994, Koehn and Monz 2006, Pang and Lee 2008, Rajpurkar et al 2016]

[Williams 1992, Lafferty et al 2001, Collins 2002, Roth and Yih 2005, Vinyals et al 2015]

12

NLP as a Arms-Race

Data

- Benchmark evaluation dataset
- Paired with training dataset
- Examples:
  - Example Translation Pairs (WMT)
  - Parse Trees for Language (PennTreebank)
  - IMDB Movie Review Dataset
  - GeoQuery Dataset

[Marcus et al 1994, Koehn and Monz 2006, Pang and Lee 2008, Rajpurkar et al 2016]

Model

- Improve accuracy, efficiency, interpretability
- Examples:
  - Conditional Random Field (CRF)
  - Integer Linear Programming (ILP)
  - Seq-to-Seq Model (RNN)
  - Reinforcement Learning

[Williams 1992, Lafferty et al 2001, Collins 2002, Roth and Yih 2005, Vinyals et al 2015]

Zero-th step: Find topics

- What topics in NLP are you interested in?
- Could be application focused — summarization, conversation agent, question answering, negotiation, machine translation, image captioning, instruction following, information extraction, coreference resolution, entity linking, entity typing,
- Could be method driven — efficient inference, generative model, unsupervised learning, domain adaptation, etc…
- Try to make your “topic” more specific — what aspect of the application you would like to focus on?
- Look at the first lecture slides to get some ideas!

Types of tasks in NLP

- Emulation Task:
  - Models emulating human intelligence
  - Many tasks in NLP fall under this category, model estimating human annotations

- Discovery Task:
  - Uncovering patterns in language that humans may not recognize
  - Predicting popularity of tweets, memorability of movie quotes, detecting fake reviews, etc

Areas in NLP

SUBMISSIONS TOPICS

ACL 2022 aims to have a broad technical program. Relevant topics for the conference following areas (in alphabetical order):

- Computational Social Science and Cultural Analytics
- Dialogue and Interactive Systems
- Discourse and Pragmatics
- Ethics and NLP
- Information Extraction
- Information Retrieval and Text Mining
- Interpretability and Analysis of Models for NLP
- Language Grounding to Vision, Robotics and Beyond
- Linguistic Theories, Cognitive Modeling, and Psycholinguistics
- Machine Learning for NLP
- Machine Translation and Multilinguality
- NLP Applications
- Phonology, Morphology, and Word Segmentation
- Question Answering
- Resources and Evaluation
- Semantics: Lexical
- Semantics: Sentence-level Semantics, Textual Inference, and Other Areas
- Sentiment Analysis, Stylistic Analysis, and Argument Mining
- Speech and Multimodality
- Summarization
- Syntax: Tagging, Chunking and Parsing
- Theme: "Language Diversity: from Low-Resource to Endangered Languages"

First Step: Literature Review

- Start early, and look wide and deep!
- You do not want to re-invent the wheels.
- You will learn about common tricks, libraries, etc that will make your life easier.

Literature Review

- 1. Do a keyword search on Google Scholar, Semantic Scholar, or ACL anthology
- 2. Download papers that seem relevant
- 3. Skim the abstracts, introduction, and previous work section
- 4. Identify papers that look relevant, appear often, and have lots of citations
- 5. Download those papers (and go back to step 3)

Where to find good papers

- Where to find the most trustworthy papers:
  - NLP: Proceedings of ACL conferences (ACL, NAACL, EACL, EMNLP, CoNLL, LREC), Journal of Computational Linguistics, TACL, COLING, arXiv*
  - Machine Learning/AI: Proceedings of NeurIPS, ICML, ICLR, AAAI, IJCAI, and arXiv*
  - Computational Linguistics: Journals like Linguistic Inquiry, NLLT, Semantics and Pragmatics

ACL Anthology FAQ Corrections Submissions Search...

Welcome to the ACL Anthology!

The ACL Anthology currently hosts 71273 papers on the study of computational linguistics and natural language processing.

Subscribe to the mailing list to receive announcements and updates to the Anthology.

The Anthology can archive your poster or presentation! Please submit them in PDF format by **filling out this form**. Attachments will be distributed under the terms of the **CC-BY-4.0 license**.

ACL Events

Venue	2021 – 2020	2019 – 2010	2009 – 2000	1999 – 1990	1989 and older
AACL	20				
ACL	21 20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 87 86 85 84 83 82 81 80 79
ANLP			00	97 94 92	88 83
CL	20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 87 86 85 84 83 82 81 80 78 77 76 75 74
CoNLL	20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97	
EACL	21	17 14 12	09 06 03	99 97 95 93 91	89 87 85 83
EMNLP	20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96	
Findings	21 20				
NAACL	21	19 18 16 15 13 12 10	09 07 06 04 03 01 00		
SemEval	21 20	19 18 17 16 15 14 13 12 10	07 04 01	98	
*SEM	21 20	19 18 17 16 15 14 13 12			
TACL	21 20	19 18 17 16 15 14 13			
WMT	20	19 18 17 16 15 14 13 12 11 10	09 08 07 06		
WS	20	19 18 17 16 15 14 13 12 11 10	09 08 07 06 05 04 03 02 01 00	99 98 97 96 95 94 93 92 91 90	89 88 86 84 81 79 77
SIGs	ANN | BIOMED | DAT | DIAL | EDU | EL | FSM | GEN | HAN | HUM | LEX | MEDIA | MOL | MORPHON | MT | NLL | PARSE | REP | SEM | SEMITIC | SLAV | SLPAT | SLT | TYP | UR | WAC				

NLP Conferences

- ACL
- EMNLP
- NAACL
- EACL, IJCNLP, COLING, LREC…

- “Best paper” awards — usually well written, strong papers!
- Papers covered in university seminar classes

- 9 page (long), 5 page (short) papers
- Double-blind reviews

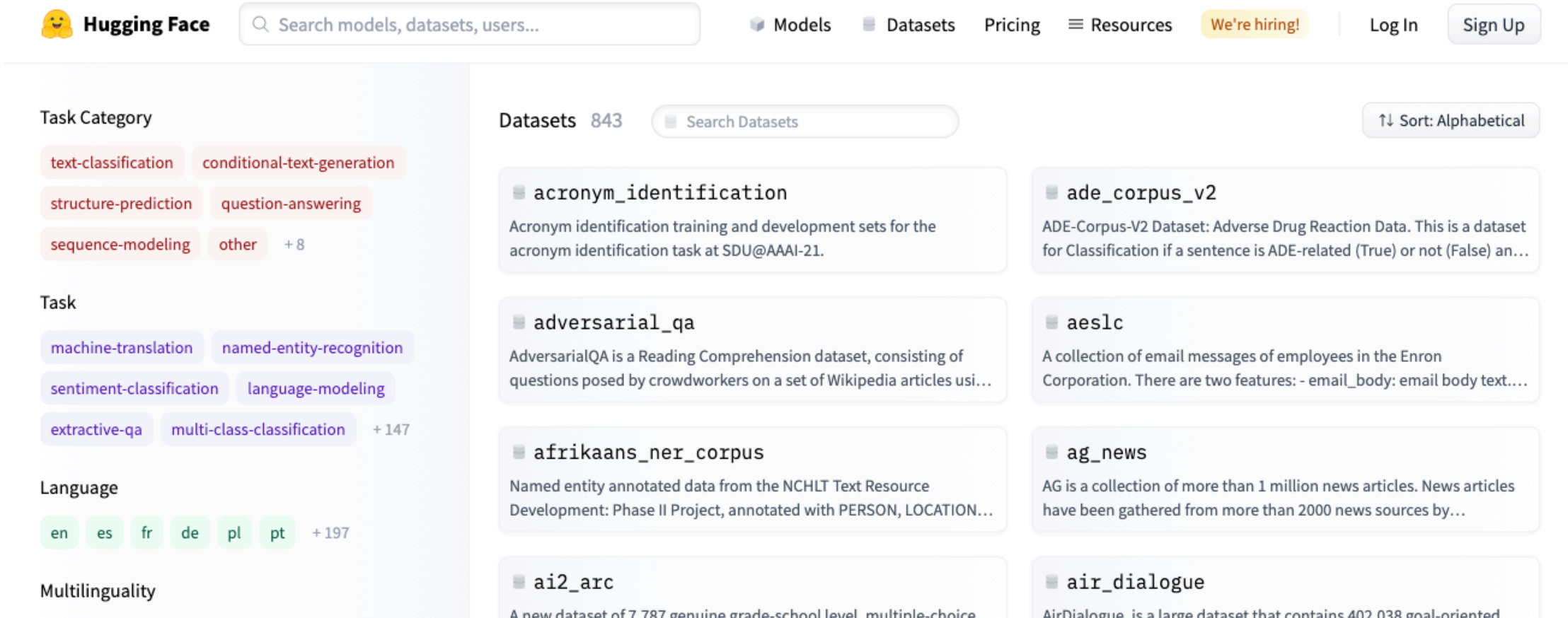
- Different types of papers: modeling, dataset, analysis, etc!

Today

- Getting started with NLP research project
  - Where should we start?
  - Dataset
  - Evaluation
  - Model
  - Brief discussion about ethics

Where to find dataset?

- Existing datasets
  - Linguistic Data Consortium
  - HuggingFace “Datasets”



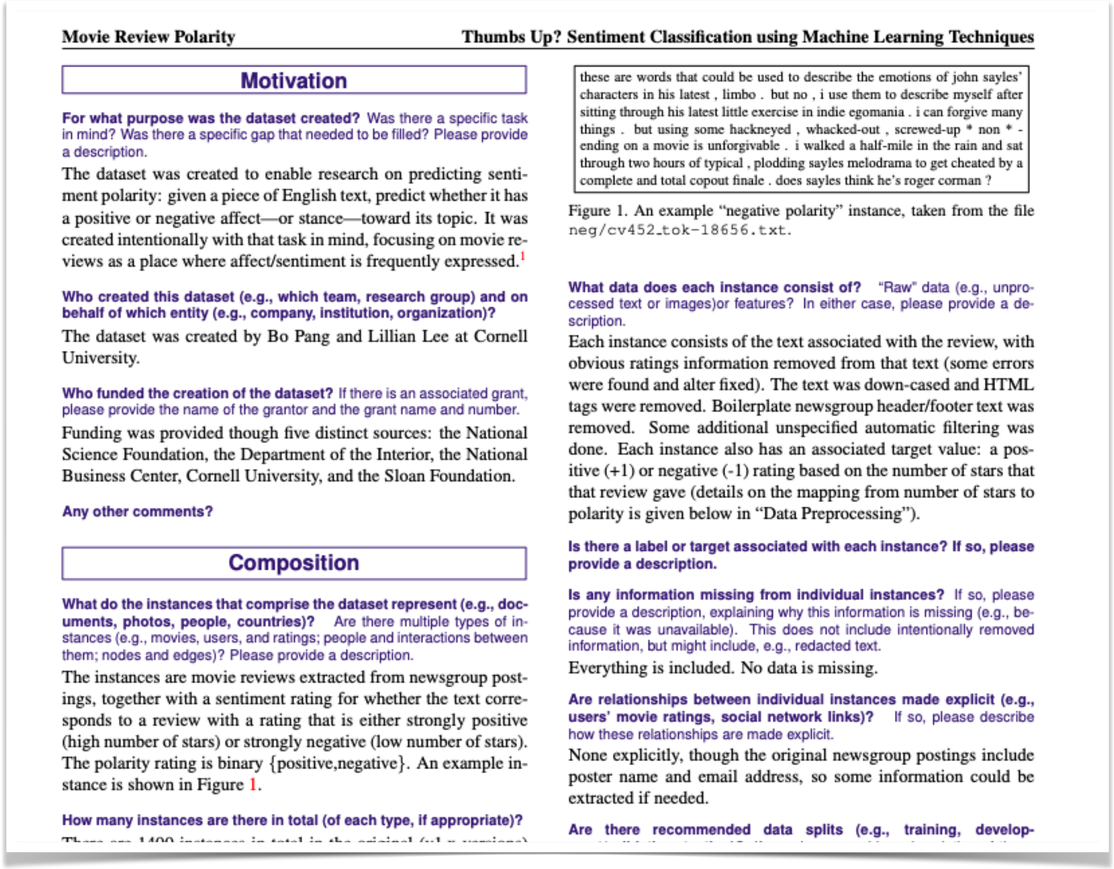
Where to find dataset?

- Existing datasets
  - Linguistic Data Consortium
- Find them in the wild
  - Example: StackOverflow, MoveiReview datasets
  - Careful with copyright
- Build them
  - Collect dataset with experts
  - Crowdsourcing
  - But be careful with generating artificial data that does not reflect real world

Dataset

- How are you selecting / curating a dataset?

  - Objective
  - Development Process
  - Collection Process
  - Uses
  - Distribution
  - Maintenance
  - Impact



Movie Review Polarity **Thumbs Up? Sentiment Classification using Machine Learning Techniques**

Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

The dataset was created to enable research on predicting sentiment polarity: given a piece of English text, predict whether it has a positive or negative affect—or stance—toward its topic. It was created intentionally with that task in mind, focusing on movie reviews as a place where affect/sentiment is frequently expressed.[1]

Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

The dataset was created by Bo Pang and Lillian Lee at Cornell University.

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

Funding was provided though five distinct sources: the National Science Foundation, the Department of the Interior, the National Business Center, Cornell University, and the Sloan Foundation.

Any other comments?

Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

The instances are movie reviews extracted from newsgroup postings, together with a sentiment rating for whether the text corresponds to a review with a rating that is either strongly positive (high number of stars) or strongly negative (low number of stars). The polarity rating is binary {positive,negative}. An example instance is shown in Figure 1.

How many instances are there in total (of each type, if appropriate)?

these are words that could be used to describe the emotions of john sayles' characters in his latest , limbo . but no , i use them to describe myself after sitting through his latest little exercise in indie egomania . i can forgive many things . but using some hackneyed , whacked-out , screwed-up * non * - ending on a movie is unforgivable . i walked a half-mile in the rain and sat through two hours of typical , plodding sayles melodrama to get cheated by a complete and total copout finale . does sayles think he's roger corman ?

Figure 1. An example "negative polarity" instance, taken from the file `neg/cv452_tok-18656.txt`.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images)or features? In either case, please provide a description.

Each instance consists of the text associated with the review, with obvious ratings information removed from that text (some errors were found and alter fixed). The text was down-cased and HTML tags were removed. Boilerplate newsgroup header/footer text was removed. Some additional unspecified automatic filtering was done. Each instance also has an associated target value: a positive (+1) or negative (-1) rating based on the number of stars that that review gave (details on the mapping from number of stars to polarity is given below in "Data Preprocessing").

Is there a label or target associated with each instance? If so, please provide a description.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

Everything is included. No data is missing.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.

None explicitly, though the original newsgroup postings include poster name and email address, so some information could be extracted if needed.

Are there recommended data splits (e.g., training, develop-

Datasheets for datasets [Gebru et al, 2018]

Dataset: Exclusion

- Most of our annotated data is English data, especially newswire
- What about:

  Other dialects of English?

  Other languages? (Especially non-European/CJK)

  Codeswitching? (Sentences which mixes multiple languages?
- If important technological tools don't work for some users, where does that leave those users?

See *Blodgett and O'Connor '17* for more discussions

Dataset

- Likely you will split your data into training set / development set / test set.
  - Typically random split
  - But if you are testing domain transfer, etc, you can do some other split.

- Many datasets will come with pre-defined splits — you should follow them if you'd like to compare it to existing work

- Don't plan on running multiple times on test data. (Supposed to be ran only once or twice!)

Today

- Getting started with NLP research project
  - Where should we start?
  - Dataset
  - Evaluation
  - Model
  - Brief discussion about ethics

Quantitative Evaluation

- Follow prior work, use existing metrics
- Start with simple method, use ablations to study the effectiveness of your choices
- Consider human evaluation if possible, especially for generation task
- Test statistical significance when differences are small and data is small

Quantitative Evaluation

- Break down your performance number by various metrics!
  - Popular vs. rare entities?
  - Long vs. short sentences?
- Go beyond a single number comparison:
  - Is one method better than another in terms of precision, but not in recall?
  - Are they making mistakes on the same sets of examples?
- Reporting negative results can be useful too
  - But be extra careful with making strong statements — if it didn't work, is it because of potential bugs? Wrong hyperparameters?

Qualitative Analysis

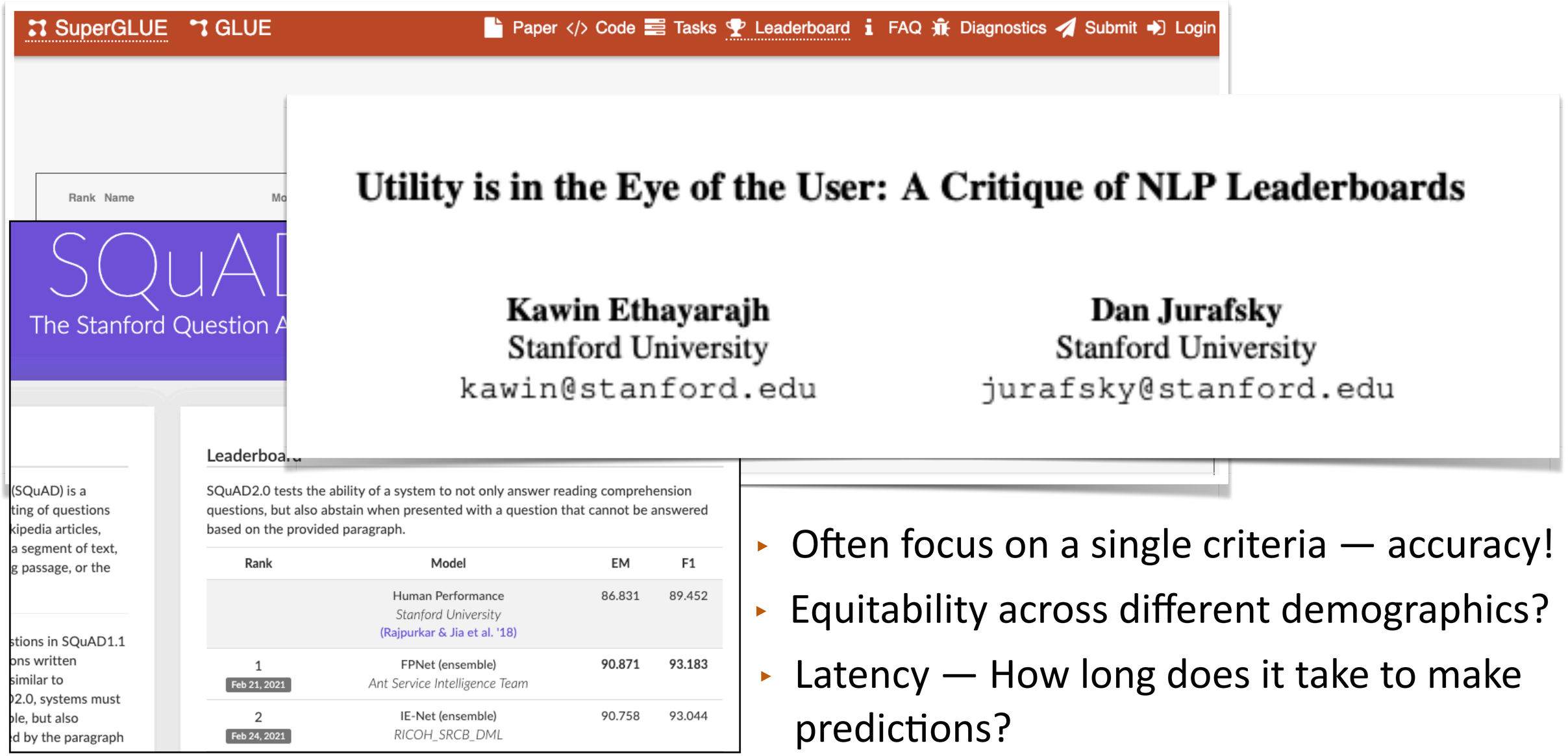
- Goal: provide evidence that your hypothesis is correct
- Often hard to evaluate with quantitative metrics:
  - e.g., Attention-based NMT models can learn the same kinds of alignments as phrase-based MT systems, but generalize better to unfamiliar words and phrases.
- Qualitative evidence can help
  - Be careful to not just cherry pick!
  - Take random samples, and categorize errors and count them
  - Visualize your model embeddings / etc with t-SNE
  - Build demo if you can!

Formative vs. Summative Evaluation

When the cook tastes the soup, that's formative;
when the customer tastes the soup, that's summative

- Formative evaluation:
  - Sanity check
  - Typically lightweight automatic metrics
  - For tuning hyperparameters, etc
- Summative evaluation:
  - Comparing your method to previous methods
  - Compare major components of your method
  - Human evaluations

NLP Leaderboards



- Often focus on a single criteria — accuracy!
- Equitability across different demographics?
- Latency — How long does it take to make predictions?

Today

- Getting started with NLP research project
  - Where should we start?
  - Dataset
  - Evaluation
  - Model
  - Brief discussion about ethics

Model

- Build a simple baseline
  - e.g., Majority class label
- Build a strong baseline
  - Existing published work can be a good baseline
  - You don’t necessarily have to beat them, especially if they are using a lot of resources that you do not have access to
- Motivate your model
  - In what aspect your proposed model improve upon baseline?

Hyperparameter Tuning

- You should tune both your baseline **AND** your new model

- During literature review, pay attention to what hyper parameters matter, and what are typical values

Focus on Real Problems

- NLP for Social Good:
  - Low resource NLP
  - Applications to help marginalized groups
  - Offer psychological help
  - Medical applications

- “Nothing about us without us” (popularized by disability rights activists Michael Masutha, William Rowland, and later James Charlton)
  - If you are building assistive technology for disabled community, you should engage with the disabled community!

How to move forward

- Hal Daume III: Proposed code of ethics
  https://nlpers.blogspot.com/2016/12/should-nlp-and-ml-communities-have-code.html
  - Many other points, but these are relevant:
    - Contribute to society and human well-being, and minimize negative consequences of computing systems
    - Make reasonable effort to prevent misinterpretation of results
    - Make decisions consistent with safety, health, and welfare of public
    - Improve understanding of technology, its applications, and its potential consequences (pos and neg)
- Value-sensitive design: vsdesign.org
  - Account for human values in the design process: understand *whose* values matter here, analyze how technology impacts those values

Global NLP community

- Universities
  - Many universities now have multiple NLP faculty members, distributed across linguistics, computer science, information science, and sometimes electrical engineering (speech)
- Companies
  - Google, Facebook, Amazon
  - Language Weaver (MT), Nuance (speech)
  - Many start-ups on different niche markets — legal, summarization,

Online NLP Community

- Many researchers have twitter accounts
  - Both individuals
  - But also organizations (Google Research), conferences
- Podcasts
  - NLP highlight
- Blogs, newsletters