

CS 6740/IS 6300: Lecture 15 - "Breaking" data/evaluation

Today's outline:

1. What is the purpose of data? One reason is evaluation, as [the Penn Treebank paper](#) said. I mention the PTB because in the "old days", it was in some sense the canonical dataset. Here's [an example results table](#) (EMNLP 2011):
2. Maybe we just need new data all the time? [Jacob Eisenstein tweet, Jun 1 2015](#)
3. Recall the landscape: new data introduced, then "solved" or "broken".
4. Demos for two NLP tasks (we can try to break the *algorithm* in class)
 1. [Sentiment analysis demo](#) at AllenNLP. Task considered in the "Build it Break it" data, and so will probably be an option for A3, since there's "regular" training data and "challenge" test sets.
 2. [Textual Entailment demo](#) at AllenNLP. Task considered in the "breaking" Levy et al. 2015 paper
5. An early example "break" of a dataset: Omer Levy, Steffen Remus, Chris Biemann, Ido Dagan. 2015. [Do supervised distributional methods really learn lexical inference relations?](#). NAACL, 970-976.

Write your name here:

Write down three sentences you'd like to try the sentiment demo on.

1:

2:

3:

