# Deep Active Learning For Named Entity Recognization

## Introduction

- We carry out incremental active learning, during the training process, and are able to nearly match state-of-the-art performance with just 25% of the original training data
- The most common approach is uncertainty sampling, in which the model preferentially selects examples for which it's current prediction is least confident
- **we mix newly annotated samples with the older ones, and update our neural network weights for a small number of epochs, before querying for labels in a new round**
- We introduce a simple uncertainty-based heuristic for active learning with sequence tagging. Our model selects those sentences for which the length-normalized log probability of the current prediction is the lowest
    - 上面这段可以说是本文最最最核心的方法了
    - We introduce a simple uncertainty-based heuristic for active learning with sequence tagging. Our model selects those sentences for which the length-normalized log probability of the current prediction is the lowest. Our experiments with the Onto-Notes 5.0 English and Chinese datasets demonstrate results comparable to the Bayesian active learning by disagreement method (Gal et al., 2017). Moreover our heuristic is faster to compute since it does not require multiple forward passes. On the OntoNotes-5.0 English dataset, our approach matches 99% of the F1 score achieved by the best deep models trained in a standard, supervised fashion despite using only a 24.9% of the data. On the OntoNotes-5.0 Chinese dataset, we match 99% performance with only 30.1% of the data. Thus, we are able to achieve state of art performance with drastically lower number of samples.

## Active Learning

- 实际的采样策略
    - **Least Confifidence (LC)**
        - Culotta & McCallum (2005) proposed to sort examples in ascending order according to the probability assigned by the model to the most likely sequence of tags
        - $$1-\max_{y_1,\ldots,y_n} \mathbb{P}\left[y_1,\ldots,y_n \mid \{\mathbf{x}_{ij}\}\right].$$
    - **Maximum Normalized Log-Probability (MNLP)**
        - $$\max_{y_1,\ldots,y_n} \frac{1}{n}\sum_{i=1}^{n}\log\mathbb{P}\left[y_i \mid y_1,\ldots,y_{n-1},\{\mathbf{x}_{ij}\}\right].$$