Submodularity-Inspired Data Selection for Goal-Oriented Chatbot Training Based on Sentence Embeddings

Mladen Dimovski[1], Claudiu Musat[2], Vladimir Ilievski[1], Andreea Hossmann[2], Michael Baeriswyl[2]

[1]IC School, EPFL, Switzerland [2]Artificial Intelligence Group, Swisscom

Slot Filling

Are there any French restaurants in downtown Toronto ?

BiLSTM model

- Task : tag the relevant keywords in the user query (sequence-tagging problem)

O	O	O	B-Cuisine
Are	there	any	French
O	O	B-Location	I-Location
restaurants	in	downtown	Toronto

Data Availability

- Existing BiLSTM models perform reasonably well if given enough training data
- What if we can afford to label only small amount of data ?

MAIN IDEA

Rank the raw unlabeled sentences according to their usefulness and select only the best ones for labeling

BAD CHOICE

I want to eat sushi !
I want to eat pizza !

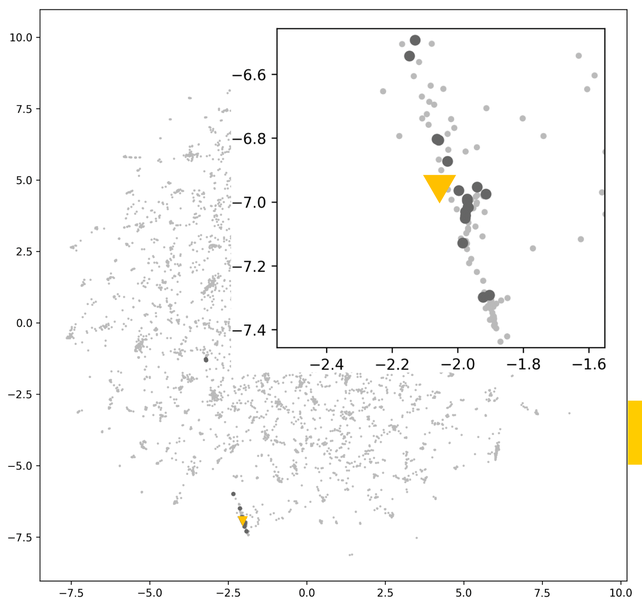
GOOD CHOICE

Can you propose a good restaurant serving beef in the city center ?
What is the best rated chinese restaurant in Lausanne ?

Contributions

1. We show that the space of raw, unlabeled sentences contains information that we can use to choose the sentences to label
2. We create a submodularity-inspired ranking function, the ratio-penalty marginal gain, for selecting the most useful sentences to label
3. We apply this data selection method to the problem of slot filling and prove that the model's performance can be considerably better with training samples chosen in an intelligent way

Sentence embeddings and sentence similarity



- Use a recently developed technique, sent2vec, that produces continuous vector representations of sentences
- Define the similarity between two sentences x and y as :
  $\text{sim}(x, y) = \exp\left(-\beta \|e(x) - e(y)\|_2\right)$
  where $e(x)$ is the embedding of the sentence x and β is the inverse of the average distance between all pairs of embeddings
- Hypothesis : closeness in the embedding space is in line with the human perception of sentence similarity

What was the movie that featured Over the Rainbow ?

- Find me the movie with the song Over the Rainbow
- What movie was the song Somewhere Out There featured in ?
- What movie features the song Hakuna Matata ?

- We experiment with 3 different publicly available datasets, each containing few thousands of sentences

MIT Restaurant Dataset

MIT Movie Dataset

ATIS Dataset

- Using some selection criteria, we select only a few dozen sentences, we reveal their labels and use them to train our model. This simulates the behavior of a system that needs to be trained for a newly available domain and we refer to it as the LOW-DATA regime.
- Performance metric : F1 score on a separate test set

Data selection methods

- Use a well-chosen submodular function to evaluate the usefulness of each subset of sentences X

A function $F : 2^V \rightarrow \mathbb{R}$ is called submodular if the value of an element diminishes as the context in which it is considered grows. Formally, F is submodular if $F(s|Y) \leq F(s|X)$ for every $X \subseteq Y$ and every $s \in V \setminus Y$.

Baselines

- Random Data Selection
- Classic Active Learning
- Randomized Active Learning

Uncertainty of the model for a sentence x of k words $u(x) = \frac{1}{k}\sum_{i=1}^{k}(1 - p_\Theta(\hat{y}_{w_i}|x))$

Coverage Score

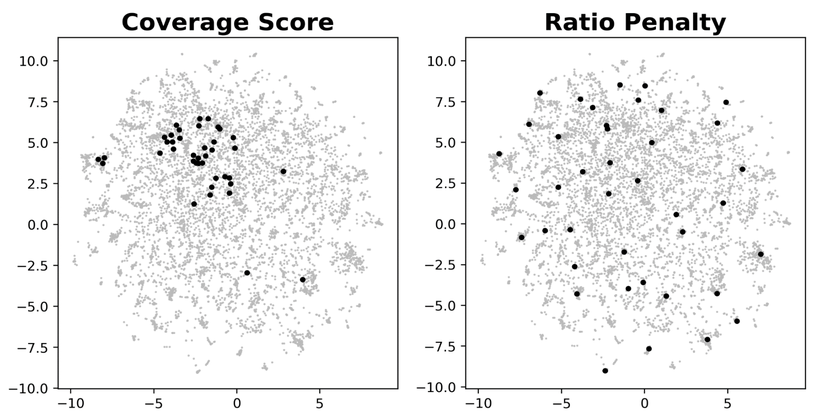
$$C(X) = \sum_{x \in X}\sum_{y \in V}\text{sim}(x, y) \qquad C(s|X) = \sum_{y \in V}\text{sim}(s, y)$$

Linear-Penalty Marginal Gain

$$D(s|X) = \sum_{y \in V}\text{sim}(s, y) - \alpha\sum_{x \in X}\text{sim}(s, x)$$

Ratio-Penalty Marginal Gain

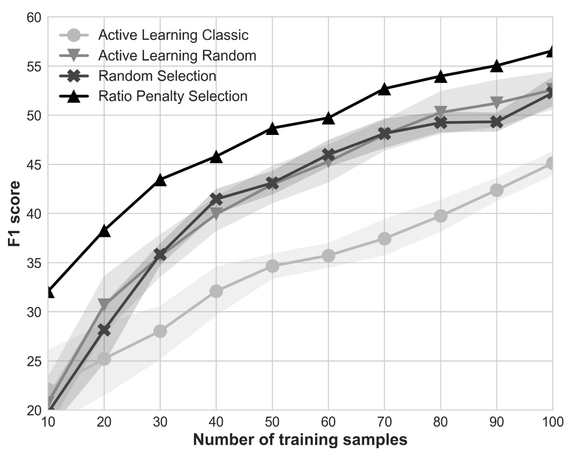
$$F(s|X) = \frac{\sum_{y \in V}\text{sim}(s, y)}{1 + \sum_{x \in X}\text{sim}(s, x)}$$



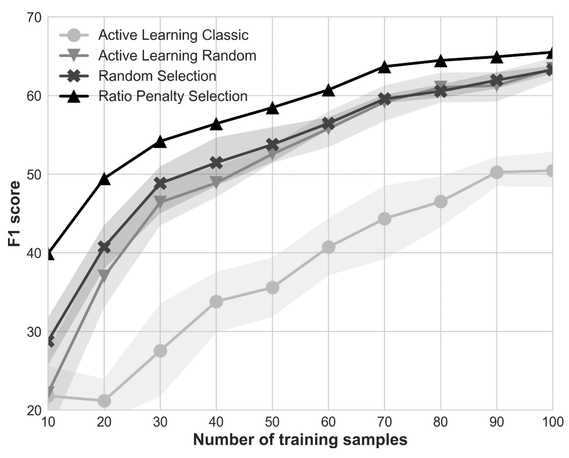
Top 3 Ratio-Penalty Sentences (MIT Restaurant Dataset)

- I need to find somewhere to eat something close by, I'm really hungry. Can you see if there's any buffet style restaurants within 5 miles of my location ?
- Find zenna noodle bar restaurant with kids friendly amenity around this place
- Where is the nearest 5 star restaurant that serves Italian food ?

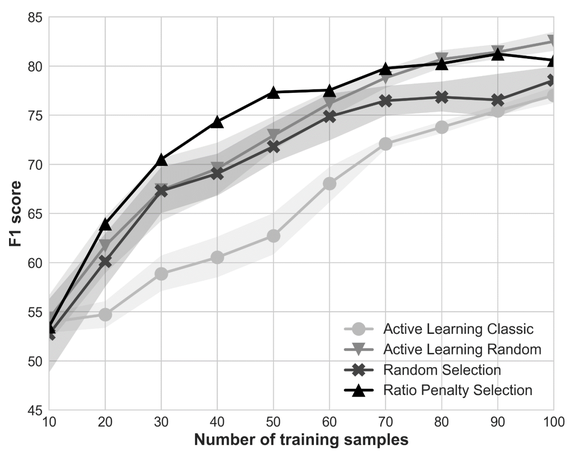
Results



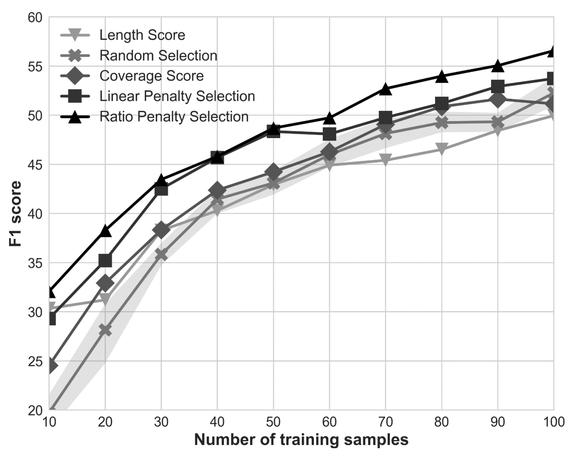
MIT Restaurant Dataset



MIT Movie Dataset



ATIS Dataset



MIT Restaurant Dataset

For more information, see: M. Dimovski et al., *Submodularity-Inspired Data Selection for Goal-Oriented Chatbot Training Based on Sentence Embeddings*, IJCAI 2018