

Active Learning via Membership Query Synthesis for Semi-supervised Sentence Classification

Raphael Schumann¹ Ines Rehbein²



1

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Institute for Computational Linguistics
Heidelberg University, Germany

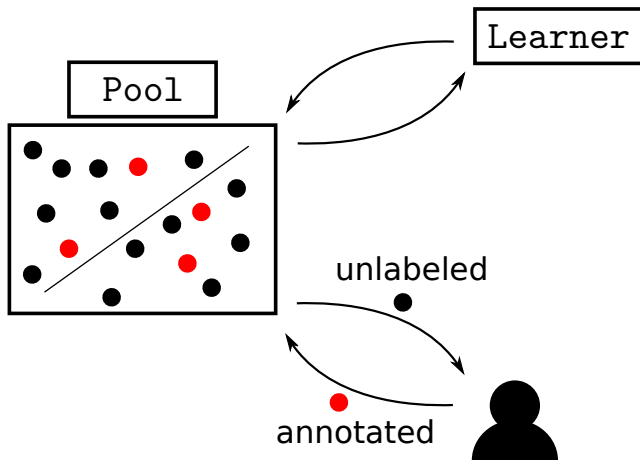
LiMo 

² Empirical Linguistics
and Computational Language Modeling
Leibniz ScienceCampus

Leibniz ScienceCampus
Heidelberg/Mannheim, Germany

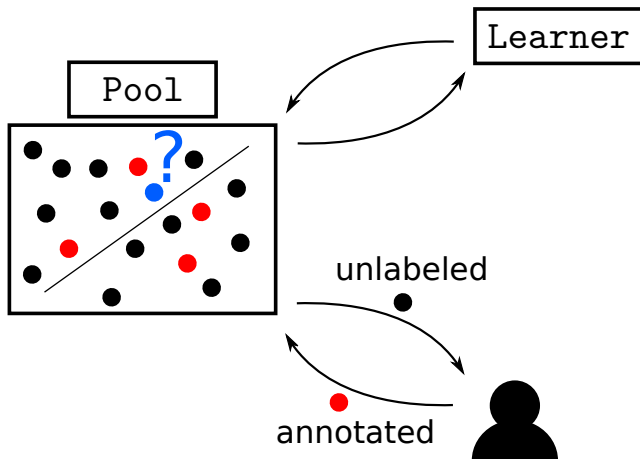
CoNLL 2019

- 1 Introduction
- 2 Method
- 3 Experiment
- 4 Analysis



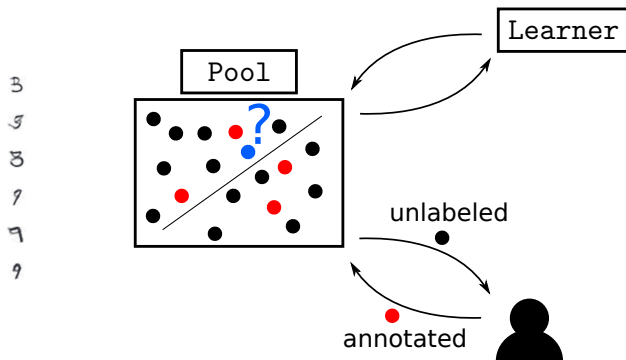
- limited to instances present in the pool

Membership Query Synthesis



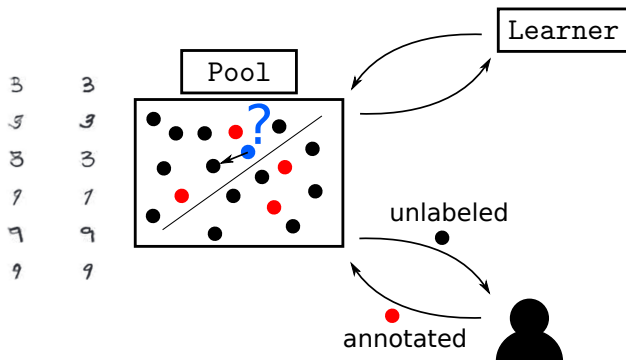
- What happens if the optimal next instance isn't in the pool

Membership Query Synthesis



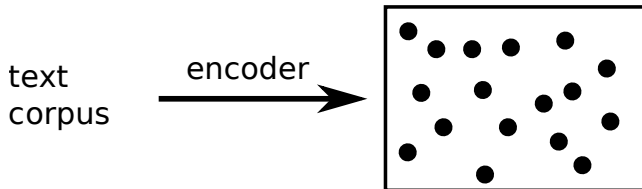
- image domain
- Wang et al. 2015

Membership Query Synthesis



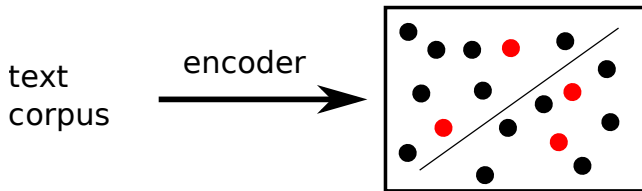
- choose nearest neighbor
- Wang et al. 2015

Semi-supervised Sentence Classification

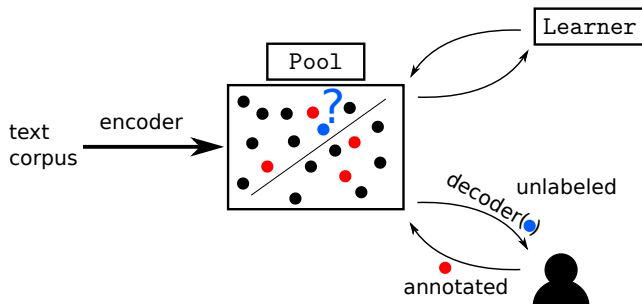


- learn continuous sentence representation

Semi-supervised Sentence Classification

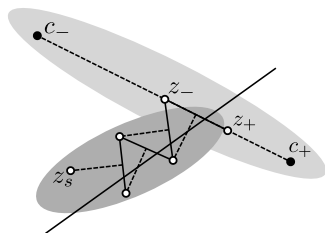


- learn continuous sentence representation
- learn classifier on labeled subset

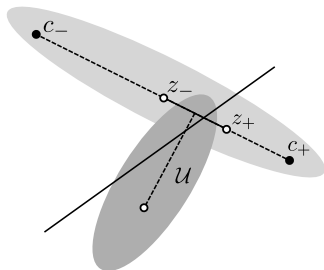


- use Variational Autoencoder (VAE) to:
 - learn continuous sentence representation
 - decode point in latent space to human readable sentence

How to find optimal next point



(a) Wang et al. 2015



(b) uniform

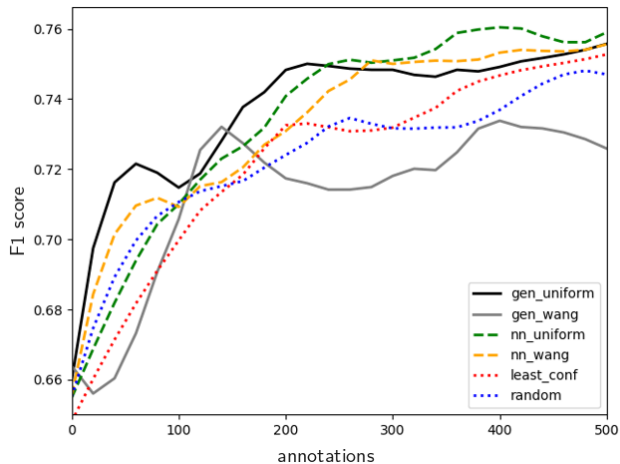
nearest neighbor decoder	nn_wang gen_wang	nn_uniform gen_uniform
other baselines	random	least confidence

	name	source	# sentences
pool	SAR14	IMDB	1.2M
test set	SST2	rottentomatoes	814

- binary sentiment classification
- only sentences up to 15 words
- overlapping sentences are removed

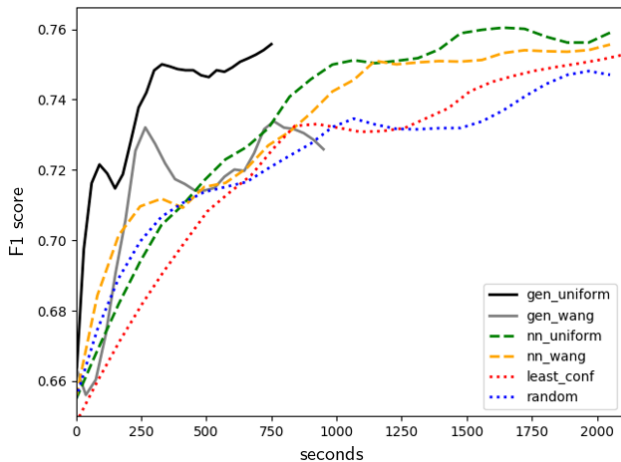
- learn classifier on minimal starting set: 'good movie' and 'bad movie'
- select next instance according to AL strategy
- annotate 500 instances selected/generated by each strategy

Results I



- average of 3 independent runs

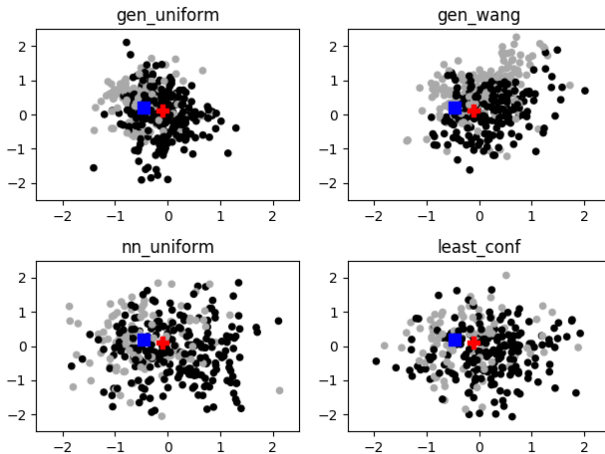
Results II



	% skips	M sec	M len	p/n
<i>gen_uniform</i>	28.1	1.4	4	1.7
<i>gen_wang</i>	20.9	1.9	5	1.2
<i>nn_uniform</i>	34.2	4.1	9	1.9
<i>nn_wang</i>	35.8	4.1	10	2.4
<i>least_conf</i>	39.0	4.2	10	2.1
<i>random</i>	36.8	4.1	9	2.6

No.	Instance	Label
1.	the acting is excellent	1
2.	powerful and moving	1
3.	this movie is very enjoyable	1
4.	a complete mess	0
5.	nothing spectacular	0
6.	absolutely terrible !	0
7.	the plot is UNK	skip
8.	well done by UNK	1
9.	the UNK is a disappointment	0

Analysis III



Thank You!

Wang, Liantao, Xuelei Hu, Bo Yuan, and Jianfeng Lu (2015). “Active learning via query synthesis and nearest neighbour search”. In: *Neurocomputing* 147. Advances in Self-Organizing Maps Subtitle of the special issue: Selected Papers from the Workshop on Self-Organizing Maps 2012 (WSOM 2012), pp. 426–434. ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2014.06.042>. URL: <http://www.sciencedirect.com/science/article/pii/S0925231214008145>.